Title

A global library screen of human primary microRNA processing utilizing a ratiometric fluorescence assay

Author

Ward, Taylor Katherine

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

A global library screen of human primary microRNA processing utilizing a ratiometric fluorescence assay

by

Taylor Katherine Ward

A dissertation submitted in partial satisfaction of the

requirements for the degree Master of Science

in Biochemistry, Molecular and Structural Biology.

2016

ABSTRACT OF THE THESIS

A global library screen of human primary microRNA processing

utilizing a ratiometric fluorescence assay

by

Taylor Katherine Ward

Master of Science in Biochemistry, Molecular and Structural Biology

University of California, Los Angeles, 2016

Professor Sriram Kosuri, Chair

MicroRNAs (miRNAs) are small, noncoding RNAs that play a vital role in gene regulation for many biological processes by binding target transcripts and marking them for degradation or post-transcriptional silencing. miRNAs undergo maturation within the nucleus beginning with cleavage of primary miRNA (pri-miRNA) by Drosha/DGCR8 to form premature miRNA (pre-miRNA). Whether this recognition relies more heavily on sequence motifs or structural features within the pri-miRNA is unknown. Algorithms based upon these potential recognition motifs predict many more miRNA sequences within the human genome than have been observed. To investigate the type of features that target a sequence for miRNA processing, we conduct a library screen of miRNA sequences within HEK293T cells to obtain a global view of miRNA processing. A reporter system consisting of GFP-P2A-puromycin and a mCherry—pri-miRNA fusion library was utilized. Library sequences that are processed by Drosha/DGCR8 lead to cleavage and degradation of the mCherry transcript and a decrease in fluorescence. These sequences are then analyzed for sequence motifs or structural features to elucidate how miRNAs are recognized and targeted for maturation.

The thesis of Taylor Katherine Ward is approved.

Guillaume Chanfreau

Feng Guo

Sriram Kosuri, Committee Chair

University of California, Los Angeles

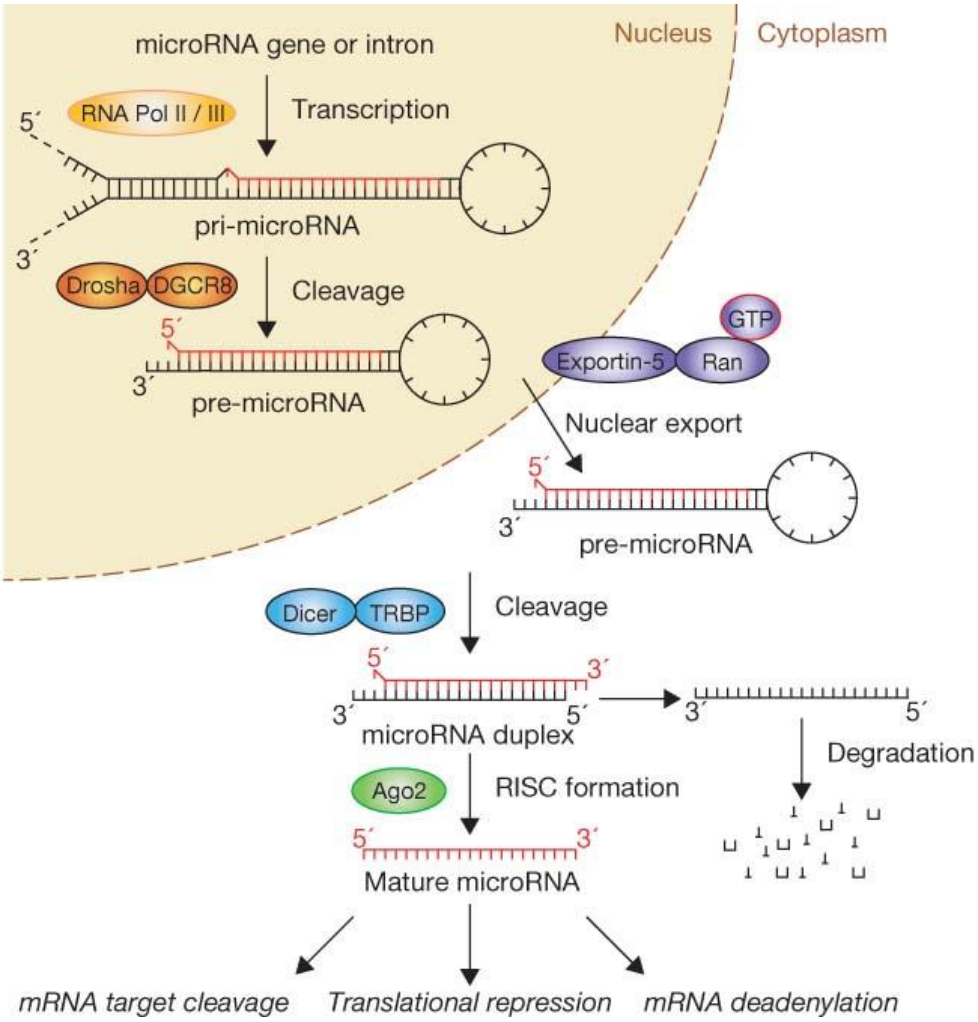
2016

Table of Contents

Introduction

MicroRNAs (miRNAs) are small, noncoding RNAs that act as guide molecules in RNA silencing and play a vital role in gene regulation for many biological processes (Ha and Kim, 2014). These small RNAs can be found in both animals and plants but differ greatly in terms of biogenesis, sequence and precursor structure, thus the focus here shall be on animal miRNAs (Miller and Waterhouse, 2005). Animal miRNAs are ~22 nucleotides in length and function in association with Argonaute (AGO) family proteins (Bartel, 2009). miRNA mediated RNA silencing occurs by the miRNA binding target transcripts and marking them for degradation or post-transcriptional silencing through the action of factors recruited by AGO proteins (Huntzinger and Izaurralde, 2011). The majority of microRNA binding sites exist within the 3' untranslated region (3' UTR) of protein coding mRNAs, although non-canonical binding

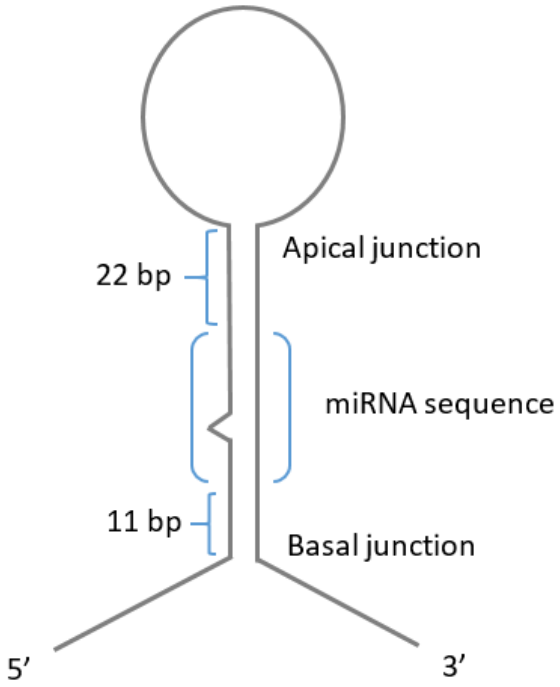
Figure 1. Biogenesis of microRNAs



Pathway of microRNA maturation: from initial transcription within the nucleus to recognition and cleavage by the Microprocessor followed by export to the cytoplasm where the final cleavage step mediated by Dicer occurs (Winter et. Al., 2009).

sites have also been shown to exist (Schnall-Levin et. Al., 2010). The exact method of silencing depends upon sequence complementarity between the conserved, 5′ seed region of the miRNA and its target sequence. Perfect base-pairing leads to AGO2 mediated cleavage of the transcript and subsequent degradation, while incomplete pairing at the 5′ seed region of the miRNA results in transcriptional repression by impeding the translational machinery through the action of any number of other AGO proteins (Saito and Sætrom, 2010). It is estimated that over 1000 miRNAs are encoded within the human genome, targeting approximately 60% of genes (Friedländer et. Al., 2014). Dysregulation of the tightly controlled biogenesis of miRNAs can prevent them from performing their vital, regulatory function and lead to health issues ranging from myocardial infarction to autoimmune disease and neurological disorders (Ardekani and Naeini, 2010), (Chang and Mendell, 2007). The importance of miRNAs in regulating many biological processes indicates that an understanding of the biological steps that lead to properly functioning, mature miRNAs is essential. miRNA genes are present within the intronic regions of both coding and non-coding transcripts, although some miRNAs are encoded within exonic regions instead (Cullen, 2004). The majority of miRNA promoters have not been studied thoroughly, but many intronic miRNAs have been shown to share the promoter of their host gene. Other miRNAs, including exonic miRNAs, have their expression driven by a separate promoter from their host, and miRNA expression can even be driven by multiple promoters (Rodriguez et. Al., 2004). RNA Polymerase II transcribes the majority of miRNAs and pri-miRNAs exist

Figure 2. Pri-miRNA cleavage



Schematic depicting the sites within the pri-miRNA sequence where the Microprocessor complex catalyzes cleavage and subsequent release of the pre-miRNA from the primary transcript.

within the resulting primary transcript as sequence elements that form an internal hairpin by undergoing intramolecular base-pairing (Lee et. Al., 2004). Upon transcription, mammalian miRNAs initially undergo several maturation steps within the nucleus beginning with recognition and cleavage of the primary miRNA (pri-miRNA) by the Drosha/DGCR8 Microprocessor complex to form premature miRNA (pre-miRNA) (Denli et. Al., 2004). The nuclear membrane protein exportin-5 exports the pre-miRNA to the cytoplasm, where it undergoes a final cleavage step by the RNase III endonuclease Dicer at the apical loop, releasing an RNA duplex of a miRNA pair (Jiang et. Al., 2005). The members of this pair will predominantly mature into the biologically active guide miRNA that makes up 95-98% of the miRNA pair population, as well as the less common passenger miRNA strand (Siomi and Siomi, 2010). The distinction between guide and passenger strand occurs when the duplex miRNAs are loaded onto the AGO protein associated pre-RNA-induced silencing complex (pre-RISC) by Dicer (Ha and Kim, 2014). After loading, the guide strand is determined based on relative thermodynamic stability of the two ends of the RNA duplex: the 5′ terminus of the guide RNA corresponds to the more unstable end. The duplex is then promptly unwound by a helicase activity and the passenger strand released, generating a mature RISC and miRNA (Gregory et. Al., 2005). The step in miRNA biogenesis which is of most interest to us is the initial recognition of the pri-miRNA by Drosha/DGCR8, whereby the complex binds the hairpin and cleaves by a measure and cut function: one cut occurs 20 nucleotides below the apical junction and the other 11 nucleotides above the basal junction [Figure 2] (Ma et. Al., 2013), (Han et. Al., 2004). The internal hairpin structure formed by pri-miRNA sequences within the initial transcript is thought to contribute to the recognition of a sequence by the Microprocessor as a pri-miRNA, but it is still uncertain as to how large of a role sequence identity plays in recognition (Kwon et. Al., 2016). The sequence elements UG and CNNC are commonly found within the basal junction where dsRNA meets ssRNA and the UGUG motif exists within the terminal loop. (Fang and Bartel, 2015). These motifs have been shown to function as binding sites for factors which enhance miRNA processing efficiency and at least 1 in 3 of these motifs

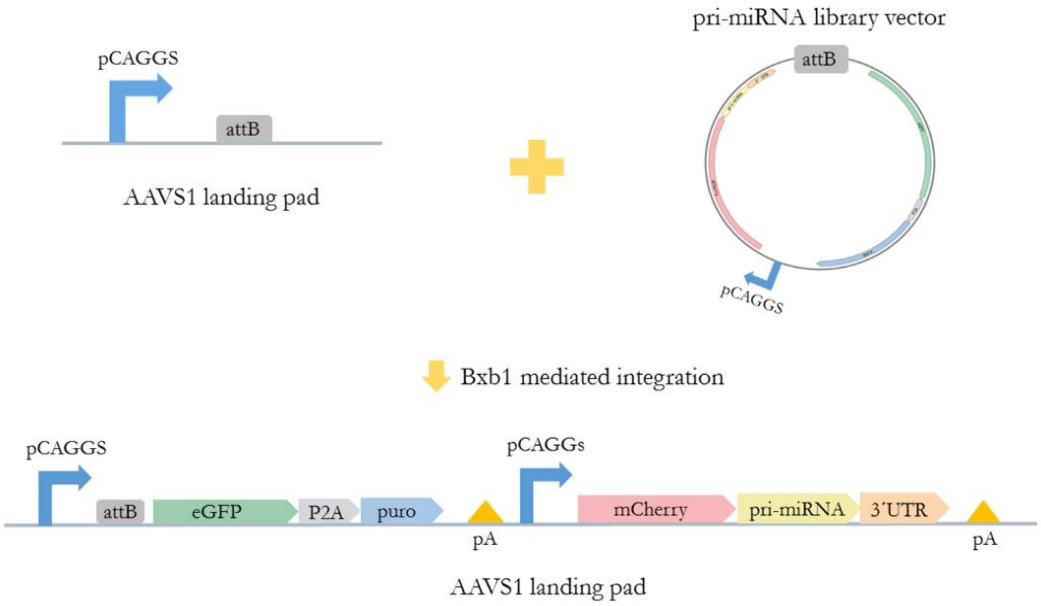
can be found in 79% of human pri-miRNAs. Studies seeking to determine whether Drosha/DGCR8 relies more heavily on sequence motifs or structural features within the pri-miRNA have yielded algorithms that predict many more miRNA sequences within the human genome than have been observed (Wong and Wang, 2014). To investigate the type of features that are recognized by Drosha/DGCR8 and that target a sequence for miRNA processing, we conducted a library screen of miRNA sequences within HEK293T cells to obtain a global view of miRNA processing. We integrated a reporter system consisting of eGFP-P2A-puromycin and mCherry—pri-miRNA fusion library into the genome at a landing pad in the AAVS1 locus [Figure 3]. Library sequences that are processed by Drosha/DGCR8 lead to cleavage and degradation of the mCherry transcript and a decrease in fluorescence. Analysis of variation in mCherry fluorescence in comparison to constant GFP fluorescence for the library sequences reveal those recognized by Drosha/DGCR8. These sequences can then be analyzed for sequence motifs or structural features to elucidate how miRNAs are recognized and targeted for maturation.

Results & Discussion

In order to gain a deeper understanding of microRNA processing through the study of pri-miRNA sequence recognition by the Microprocessor complex , we designed an assay that would provide a qualitative output allowing for the identification and sorting by flow cytometry of library members into populations that differ in processing efficiency. Drawing inspiration from a design previously utilized by the laboratory of Dr Feng Guo at UCLA, we generated a reporter constuct consisting of eGFP-2A-puromycin and mCherry —pri-miRNA fusion open reading frames (ORFs). We then integrated this reporter construct into the endogenous HEK 293T genome at a landing pad located in the AAVS1 locus by use of a Bxb1 recombinase. Integration into the genome at a landing pad allows for the eliminitaion of cell to cell variability in plasmid number—since each cell will each undergo the same number of integration events—as well as removing the possibility of integration at a vital genomic locus that could

lead to undesirable pleitropic affects [Figure 3]. The landing pad strain we used was HEK 293T strain originally obtained from the Ron Weiss lab, which was modified by CRISPR/Cas9 to remove a YFP locus from the genome (Duportet et. Al., 2014). We used a promoterless version of eGFP-2A-puromycin to allow for a method of selection for integration events, since the landing pad design incorporated a CAGGS promoter sequence upstream of the attB integration site which drove expression of GFP and puromycin post integration. Constitutive expression of the integrated reporter construct produces eGFP-2A-puro and mCherry—pri-miRNA fusion mRNA transcripts which then undergo translation [Figure 4]. The eGFP containing transcript produces a constant amount of mature fluorescent eGFP protein per cell, while the level of fluorescent mCherry protein varies depending upon the identity of the pri-miRNA sequence present in the 3′ UTR library cloning site. A pri-miRNA sequence which is recognized and processed more efficiently will display increased cleavage and degradation of the mCherry mRNA and subsequently lower levels of mCherry protein and fluorescence. A measure of pri-miRNA sequence recognition and processing efficiency can therefore be obtained by comparative analysis of constant GFP fluorescence

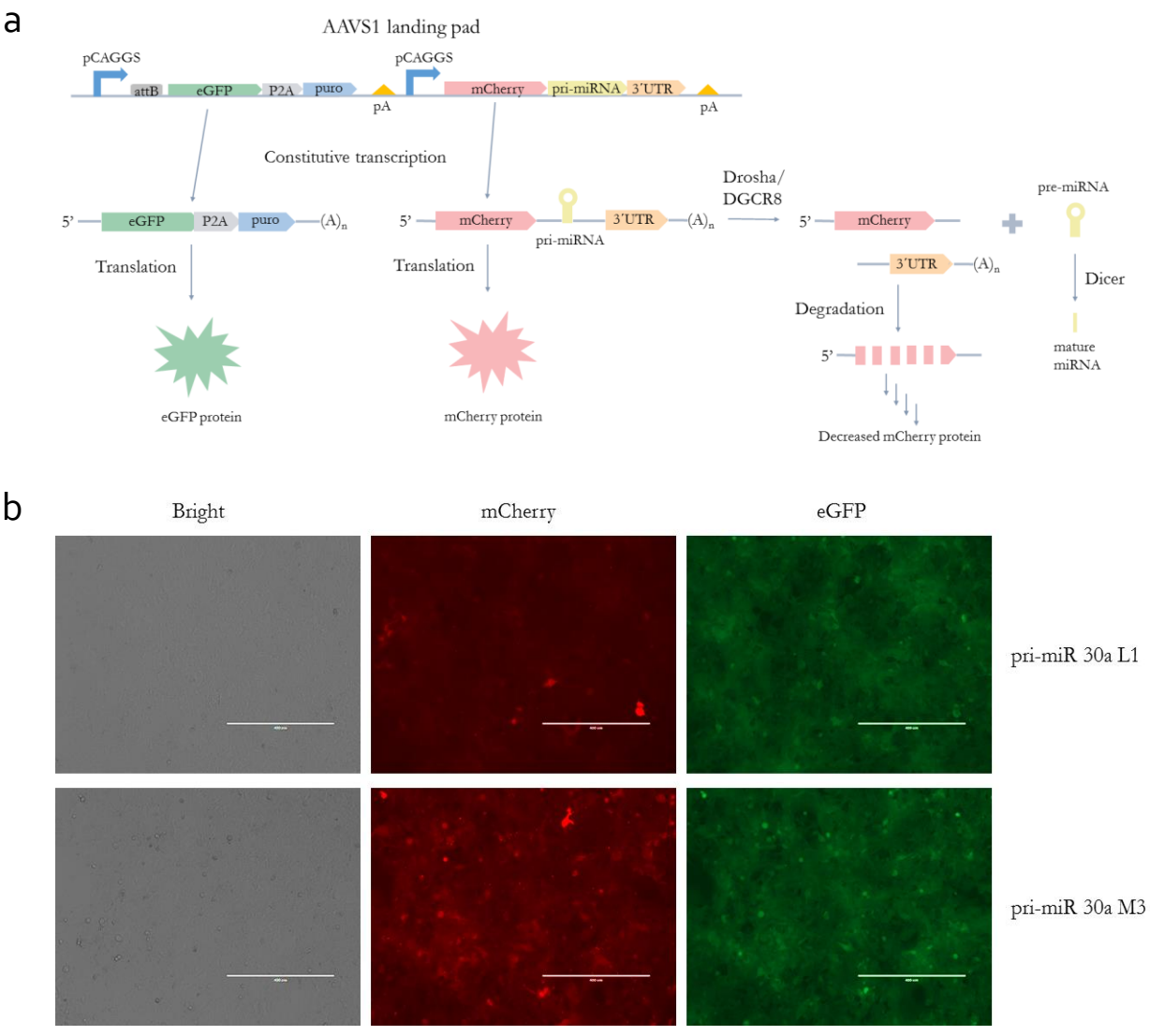
Figure 3. Genomic integration of reporter construct at AAVS1



Schematic of Bxb1 integrase mediated integration of pri-miRNA library sequences into the endogenous HEK 293T genome at the AAVS1 locus. The pre-miRNA library delivery vector was engineered with a promoterless eGFP-P2A-puro cassette to allow for puromycin selection of integration events.

with varying levels of mCherry fluorescence, with a positive correlation existing between the ratio of GFP to mCherry fluorescence and pri-miRNA processing efficiency for each library member. Initial cell

Figure 4. Ratiometric fluorescent reporter system evaluates pri-miRNA processing efficiency

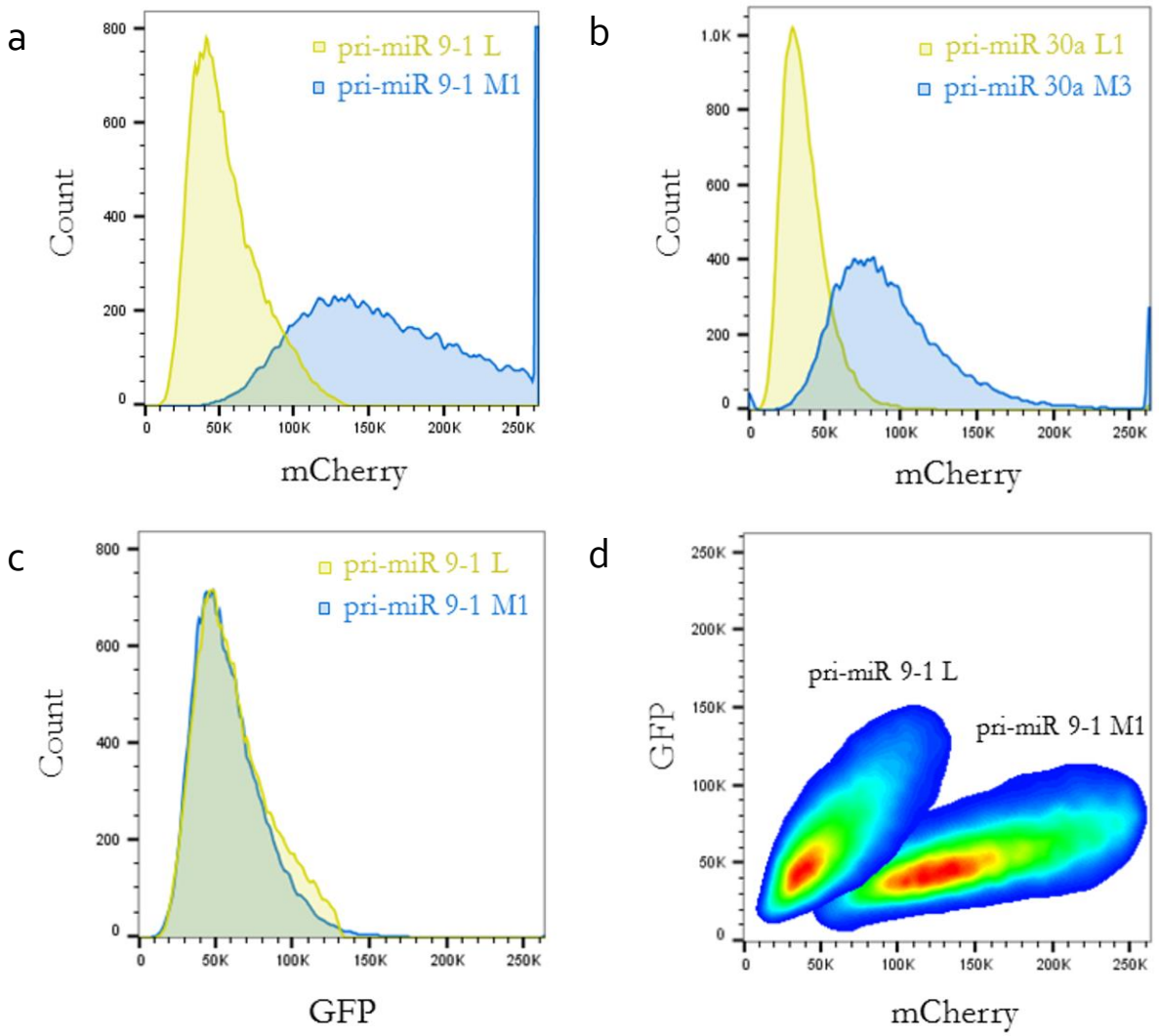


a.) Constitutive expression from reporter cassette produces a constant amount of eGFP fluorescence while the level of mCherry fluorescence depends on the pri-miRNA sequence present in the 3′ UTR. The ratio of GFP over mCherry fluorescence intensity positively correlates with pri-miRNA processing efficiency and provides a metric to sort library sequences.

b.) Cell microscopy of 293Ts with the pri-miRNA reporter construct integrated at the AAVS1 genomic locus post selection. Bright field, GFP, and mCherry fluorescence were imaged for constructs containing the pri-miR 30a (wild type) and pri-miR 30a M3 (null mutant) sequences respectively. Sequence identity had no impact on the level of eGFP fluorescence, while a significant difference in mCherry fluorescence is visible with the null mutant displaying little to no processing and mCherry cleavage and degradation in comparison to the wild type pri-miRNA.

microscopy data verified the fidelity of our reporter construct by providing qualitative data for changes in eGFP and mCherry fluorescence between constructs containing the pri-miR 30a (wild type) and pri-

Figure 5. Flow cytometry analysis of pri-miRNA controls



a). Flow cytometry data for (wild type) pri-miR 9-1 and (null mutant) pri-miR 9-1 M1 controls. Pri-miRNA 9-1 control pair display markedly different mCherry fluorescence intensity profiles with minimal (~8%) overlap. The pri-miR 9-1 population had reduced mCherry fluorescence in comparison to the pri-miR 9-1 M1 population indicating increased processing efficiency for pri-miR 9-1.
b). Same as **Fig. 4A.,** for pri-miR 30a pair.
c). Same as **Fig. 4A.,** eGFP plotted instead of mCherry. Pri-miR 9-1 and pri-miR 9-1 M1 had identical eGFP fluorescence intensity profiles, confirming eGFP expression is independent of pri-miRNA sequence identity.
d). Flow cytometry analysis of pri-miR 9-1 control pair displayed as density plots with GFP plotted against mCherry. Pri-miR 9-1 population yields lower mCherry fluorescence for a given GFP intensity than the pri-miR 9-1 M1 population. The difference in mCherry fluorescence is marked between positive and negative populations with minimal overlap.

miR 30a M3 (null mutant) sequences respectively [Figure 4]. Pri-miR 30a M3 possesses an identical

sequence to pri-miR 30a with the exception of a four base-pair mutation of the cleavage site for Drosha/DGCR8 to prevent the mutant from being recognized or processed as a pri-miRNA. As expected, the strain with pri-miR 30a M3 exhibited little to no processing of the sequence and mCherry fluorescence was relatively high intensity in comparison to the pri-miR 30a strain, which was subject to processing resulting in decreased mCherry fluorescence. Meanwhile, eGFP fluorescence intensity served as a normalization factor for cell to cell variation in fluorescent protein expression and was unaffected by sequence identity. These promising results were further supported by flow cytometry analysis of the pri-miR 9-1 and pri-miR 30a control sequence pairs [Figure 5]. Histogram plots displaying mCherry fluorescence as a function of cell number for wild type and mutant sequences for both pri-miR 9-1 and pri-miR 30a display markedly different fluorescence intensity profiles with minimal (~8%) overlap. As expected, the wild type sequence populations for each pair were shifted towards the left of the histogram where the majority of cells have reduced mCherry fluorescence—and increased microRNA processing efficiency—in comparison to the mutant sequence populations. In contrast, similar histogram displays of data which plotted GFP fluorescence intensity instead of mCherry for pri-miR 9-1 and pri-miR 9-1 M1 showed identical fluorescence intensity profiles between the wild type and mutant sequence. Most interestingly, in regards to potential application of this system to library size studies, a plot of GFP versus mCherry fluorescence for pri-miR 9-1 and pri-miR 9-1 M1 showed two distinct populations exhibiting minimal overlap. This separation of populations by high and low processing efficiency for microRNA sequences of interest suggests a potential method of analysis by gating and flow-sorting subpopulations within a larger library scale study of microRNA processing.

Materials and Methods

Cell line maintenance and modification.

HEK293T cell line was obtained from the Weiss lab and a YFP locus was removed utilizing CRIPSR/Cas9 genome editing technology. Cell lines maintained in DMEM media with 10% FBS and 5 mL

Penicillin/Strepomycin stock (100X; 10,000 IU Penicillin and 10,000 µg/ml Streptomycin) and passaged every ~3 days utilizing 0.25% TrpLE EXPRESS. Transfections carried out according to Lipofectamine 3000 protocol from Thermo Fisher Scientific. MicroRNA containing vector was co-transfected with either Bxb1 recombinase expression vector or pUC19 for negative integration control at a ratio of 4 microRNA vector to 1 Bxb1 vector or pUC19. Media was changed within 24 hours of transfection and cells were passaged ~3 days post-transfection. After 3 passages post transfection, selection for integration events was carried out by puromycin at a concentration of 2 µg/mL for at least 4 passages.

Flow cytometry

Cells prepared for flow cytometry analysis by passaging cells and resuspending at ~5x10^6 cell/mL in 500 µL of clear media such as PBS or FluoroBrite DMEM. Cell suspension was passed through filter to remove cellular aggregates into glass vial. Flow was carried out by optimizing voltages for FSC, SSC, GFP and mCherry parameters by initially flowing the following controls: unstained, single positive GFP, single positive mCherry. Data was acquired in linear mode on a BD LSRII flow cytometer and analyzed in BD FACSDIVA and FlowJo software.

References

Ardekani M.A. & Naeini M.M. The Role of MicroRNAs in Human Diseases. *Avicenna J Med Biotechnol*. (2010). 2(4): 161–179.

Bartel D. MicroRNAs: target recognition and regulatory functions. *Cell* **136** (2). (2009). 215–33

Chang TC & Mendell JT. microRNAs in vertebrate physiology and human disease. *Annu. Rev. Genomics Hum. Genet.*, 8 (2007). 215–239

Cullen RB. Transcription and Processing of Human microRNA Precursors. *Molecular Cell*. (2004) 16:861–865.

Denli A, Tops B, Plasterk R, Ketting R, Hannon G. Processing of primary microRNAs by the Microprocessor complex. *Nature*. (2004) 432:231–235.

Duportet X, Wroblewska L, Guye P, Li Y, Eyquem J, Rieders J, Rimchala T, Batt G, Weiss R. A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res*. (2014). 42, 13440–13451.

Fang W & Bartel D. The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes. *Mol Cell*. (2015). 60:131-145.

Friedländer MR, Lizano E, Houben AJ, Bezdan D, Báñez-Coronel M, Kudla G, Mateu-Huertas E, Kagerbauer B, González J, Chen KC, LeProust EM, Martí E, Estivill X. Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biology*. (2014). Apr; 15:R57.

Gregory R, Chendrimada T, Cooch N, Shiekhattar R. Human RISC Couples MicroRNA Biogenesis and Posttranscriptional Gene Silencing. *Cell*. (2005). Nov; 123(4): 631–640.

Ha M & Kim V. Regulation of microRNA biogenesis. *Nature Reviews*. (2014). 15: 509-522.

Huntzinger E & Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews Genetics*. (2011). Feb; 12: 99-110.

Jiang F, Ye X, Liu X, Fincher L, McKearin D, Liu Q. Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. *Genes Dev*. (2005). 19: 1674–1679

Han JF, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Dev*. (2004). 18: 3016-3027

Kwon C, Nguyen TA, Choi YG, Jo MH, Hohng S, Kim VN, Woo JS. Structure of Human DROSHA. *Cell*. (2016). Jan;164(1-2): 81–90.

Lee Y. et. Al. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal*. (2004). 23, 4051-4060.

Ma H, Wu Y, Choi JG, Wu H. Lower and upper stem–single-stranded RNA junctions together determine the Drosha cleavage site. *PNAS*. (2013) Jun;110(51): 20687–20692.

Miller A & Waterhouse P. Plant and animal microRNAs: similarities and differences. *Funct Integr Genomics*. (2005) Jul;5(3):129-35.

Rodriguez A, Griffiths-Jones S, Ashurst LJ, Bradley A. Identification of Mammalian microRNA Host Genes and Transcription Units. Genome Research. (2004). 14:1902–1910.

Saito T & Sætrom P. MicroRNAs – targeting and target prediction. *Elsevier*. (2010). Jul;27(3): 243-49.

Schnall-Levin M, Zhao Y, Perrimon N, Berger B. Conserved microRNA targeting in *Drosophila* is as widespread in coding regions as in 3'UTRs. *PNAS*. (2010) Jul;107(36): 15751–15756.

Siomi H & Siomi M. Posttranscriptional Regulation of MicroRNA Biogenesis in Animals. *Molecular Cell*. (2010). May;38(3): 323–332.

Winter J, Jung S, Keller S, Gregory R, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature Cell Biology.* (2009). 11, 228 – 234.

Wong N & Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res*. (2014).