The Thesis Committee for Rachel Priya Chirayil
Certifies that this is the approved version of the following thesis:

IMPACT OF NON-CODING RNAS IN GENETIC DISEASE AND PAPILLOMAVIRUS LIFECYCLE

APPROVED BY

SUPERVISING COMMITTEE:

Supervisor:

Christopher S. Sullivan

Jason Upton

IMPACT OF NON-CODING RNAS IN GENETIC DISEASE AND PAPILLOMAVIRUS LIFECYCLE

by

Rachel Chirayil

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

August 2017

Dedication

I would like to dedicate this thesis to my mother, my first inspiration to study science and my first role model, personally and professionally. Without your love and support, as well as your tireless work ethic and sacrifices, I would know what it means to be a strong and independent scientist. From the first lesson in my graduate studies at UT, that I would have to work my butt out, to my last, that things aren't always fair, and what defines us is how we move forward from monumental setbacks, you have been there by my side for every step of the way. Any good that I do in the sphere of science, and in the world, I do because you led the way, and I will forever be grateful for that.

Acknowledgements

I would like to thank all of the members of the Sullivan lab for their support and guidance. From fielding endless and inane questions to taking Topo Chico or coffee breaks with me, you have all helped me in ways big and small to get through my graduate studies.

I would also like to thank all members of the Upton lab for their support, guidance, and thoughtful critiques, as well as use of their space reagents and thanks to Dr. Jason Upton for reading my thesis.

I would finally like to thank all of my family and friends for their support during my graduate studies. I would especially like to thank my parents and siblings, who put up will all of my insane hours and missed holidays, and still manage to love and cherish me. I cannot thank you enough.

The work reported in this thesis will be submitted for publication in the near future. A significant portion of the writing and some experimentation in Chapter 3 was contributed by lab members and collaborators. All work and writing in Chapter 2 is sole work of the author.

Abstract

IMPACT OF NON-CODING RNAS IN GENETIC DISEASE AND PAPILLOMAVIRUS LIFECYCLE

Rachel Priya Chirayil, M.A .

The University of Texas at Austin, 2017

Supervisor: Christopher S. Sullivan

Non-coding RNAs are vital to several diverse biological processes. Untranslated regions of mRNAs (UTRs) and variability therein can often cause differential regulation of transcripts. This regulation is often effected by microRNAs (miRNAs), small non-coding RNAs that interacts with 3'UTRs to regulate gene expression post-transcriptionally. Here we investigate the variation in the 3'UTR of SERPINA1 gene, mutations of which cause the often life-threatening Alpha-1-antitrypsin deficiency (A1AD), and effects on miRNA. We screened 50 miRNA mimics and determined their overexpression does not affect secreted alpha-1-antitrypsin (AAT). Additionally, we showed through the use of a poxvirus protein to degrade miRNAs that underexpression of miRNAs does not affect secreted AAT. Using luciferase reporters we determine that in cultured cells, there are no liver-specific trans-factors that act on the SERPINA1 3'UTR, of any length, to

regulate gene expression. We sequenced the 3'UTRs from the largest cohort of of A1AD patients to date and determined that 3'UTR variation does not contribute to disease severity in patients with one form of A1AD, and that in a cell culture context, miRNAs do not regulate secreted AAT. We conclude that neither 3'UTR variation nor miRNAs affect wild-type secreted AAT.

In addition to genetic disease, miRNAs regulate multiple aspects of the host-pathogen interface. Until now no widely accepted PV-encoded miRNAs have been described. We have developed miRNA Discovery by forced Genomic Expression (miDGE), a new wet bench approach to miRNA identification that screens numerous pathogen genomes in parallel. Using miDGE, we screened over 75 different PV genomes for the ability to code for miRNAs. We conclusively demonstrate a lack of PV miRNA expression in cancers associated with infections of several high risk HPVs. However, we identified five different miRNAs encoded by four different PVs (Human PVs 17, 37, 41 and Fringilla colebs (Fc) PV). We show that miRNAs from two PVs (HPV41 & FcPV) are able to regulate transcripts corresponding to the early region of the PV genome.. Combined, these findings identify the first canonical PV miRNAs and support that miRNAs of either host or viral origin are important regulators of the PV life cycle.

Table of Contents

List of Tables

List of Figures

IMPACT OF NON-CODING RNAS IN GENETIC DISEASE AND PAPILLOMAVIRUS LIFECYCLE

Chapter 1: Introduction and overview of microRNAs and Papillomaviruses

In recent years, there has been a dramatic rise in appreciation for the diverse regulatory roles of non-protein-coding RNAs. Deep sequencing technology has uncovered a surfeit of wide-ranging biological roles for non-coding RNAs (ncRNAs) in across a large variety of eukaryotic and prokaryotic genomes. Several types of ncRNA regulate gene expression, controlling mechanisms such as transcription, pre-mRNA processing, transcript turnover, translation, and nuclear export (as reviewed in 13).

RNA interference (RNAi) is the mechanism by which small non-coding RNAs, in complex with protein machinery, regulate gene expression by targeting messenger RNA (mRNA) molecules. MicroRNAs (miRNAs), small 19-22 nucleotide (nt) RNAs that target mRNAs for cleavage or repression of translation, belong to one such class of ncRNAs (13, 22, 43). In the nearly twenty five years since their discovery in *Caenorhabditis elegans (*22, 43*),* much has been learned about these ncRNAs. With the advent of next-generation sequencing technology and computational approaches, miRNAs have been uncovered in a wide variety of eukaryotic genomes (13, 36). Several human miRNAs have been shown to regulate a diverse range of cellular processes, including the innate immune

response, cell growth and apoptosis, metabolism, developmental differentiation, he inflammatory response, and cancer (reviewed in 13, 14, and 53).

MiRNAs are canonically derived from longer RNA polymerase II (pol II)-transcribed primary transcripts (pri-miRNAs) containing stem-loop structures (44). Shorter 70-100 nucleotide (nt) precursor miRNAs (pre-miRNAs) are cleaved from the pri-miRNA by an RNase III endonuclease Drosha/ DGCR8 complex (17, 41, 44, 76) and exported into the cytoplasm by Exportin-5 in associated with RAN-GTP (74). Here, pre-miRNAs are cleaved in roughly 22 nt duplexed RNA by another RNase III endonuclease, Dicer (28, 33). This duplexed RNA is composed of a passenger strand RNA and a mature RNA, the latter of which is loaded onto the RNA-induced Silencing Complex (RISC). The miRNA-RISC complex targets complementary sequences in mRNA transcripts, typically in the non-coding 3' untranslated region (3'UTR) for degradation or translational repression (35, 61).

The 3'UTR is defined as the region of an mRNA transcript immediately following the stop codon through the poly-adenosine tail (20, 65). The average human 3'UTR is about 1000 nucleotides (65) in length, though this can vary in a context-specific manner. For instance, genes implicated in cancer progression often display shorter 3'UTRs, and some highly expressed neuronal genes have 3'UTRs up to 1300 nucleotides in length (65). As 3' UTRs are generally long, most contain several potential binding sites for an array of miRNAs. Additionally, the same type of miRNA can target a diverse range of mRNA transcripts.

Watson-Crick base pairing plays a major role in binding efficiency of and subsequent processing by RISC-bound miRNAs. Typically, complementary pairing is most critical for binding between 3'UTR sequence and a "seed" sequence in the miRNA (nucleotides 2-7), and pairing with miRNA sequence upstream or downstream of the seed can contribute to binding energy (21, 4, 34). Given the highly sequence-dependent manner of miRNA binding and the availability of multiple miRNA binding sites on most 3'UTRs, variation in 3'UTR sequences can greatly affect 3'UTR-dependent regulation. One of the most common causes of 3'UTR variation, use of alternative poly-adenlyation, has been estimated to affect almost 70% of human transcripts (20). Alternative polyadenlyation can contribute to shortening or lengthening of a gene's resultant 3'UTR, and as such can provide additional miRNA binding sites, or ablate existing sites. Single nucleotide polymorphisms (SNPs) can also contribute to differential miRNA-dependent regulation (25, 54). SNPs occur on average once every 300 nts throughout the human genome. This rate can, however be misleading, as noncoding regions are less constrained for SNPs, meaning there is a higher likelihood of variation for ncRNA than for protein-coding RNA regions. As miRNA binding is highly sequence-specific, a single nucleotide change can greatly change the degree to which a transcript is post-transcriptionally regulated (as reviewed in 54). Thus, variation in the 3'UTR can have a dramatic impact on gene expression.

RNAi has been observed in plants, insects, and nematodes to be anti-viral, though its role in the mammalian antiviral response remains controversial (1). Interestingly, several diverse viruses have been shown to use virus-encoded miRNAs in a pro-viral context. The largest portion of these are DNA viruses (16, 36), but all known miRNA-encoding viruses share one important trait: they are capable of maintaining persistent infection (16). A hallmark of persistent infection is decreased viral gene expression, allowing for infection to evade innate immune surveillance. As key regulators of gene expression, miRNAs likely play a significant role in this.

Papillomaviruses (PVs), double-stranded DNA viruses, are of special interest as they are highly related to human health. HPV16 and HPV18 are known to be responsible for the vast majority of cervical cancers, as well as a large share of throat cancers and anal cancers. As yet, there is no consensus on whether or not PVs encode miRNAs. Early attempts at sequencing have revealed no PV-encoded miRNAs. However, it has been reported that computational analysis of HPV18 revealed virus-encoded miRNAs (29). Additionally, more recent reports have suggested that small RNA library sequencing from HPV16, HPV 38, and HPV61 reveal PV-encoded miRNAs in each viral genome (57, 71). These studies do not provide complete evidence for miRNA identification, as the abundance of these small RNAs was extremely low, and neither biogenesis nor RISC activity were established. However, another study using fully infectious systems to study high risk PVs reported no evidence

of viral miRNAs derived from either HPV16 or HPV18 (76). The current dearth of information on noncoding PV RNAs is due in large part to the lack of ability to study PV infection in vitro. However, as novel technologies for small RNA sequencing develop, a clearer picture will emerge as to the absence or presence of PV miRNAs, and the biological significance in this important human pathogen.

Chapter 2: SERPINA1 3'UTR variation does not contribute to Alpha-1-Antitrypsion Deficiency disease severity

Introduction: *SERPINA1 and A1AD disease*

Alpha-1-antitryspin deficiency is a genetic disease caused by decreased functional levels of the protein Alpha-1-antitrypsin, and can lead to devastating damage to respiratory health and liver function due largely to misfolding of AAT protein and accumulation of immunoreactive isoforms in the liver (23, and references therein, 27). Alpha-1-antitrypsin (AAT) is a serine protease inhibitor that functions in the inflammatory response by regulating neutrophil activity. The 52 kDa protein, which inhibits a range of proteases, is primarily produced and secreted in the liver and is present in high amounts in the plasma. AAT activity is vital for respiratory health. In the absence of AAT its primary target, neutrophil elastase, targets elastin in the lungs, leading to decreased elasticity and respiratory distress. A1AD can lead to increasingly poor respiratory health, causing or worsening diseases such as Chronic Obstructive Pulmonary Disease (COPD), asthma, and emphysema. Accumulation of misfolded protein in the liver can additionally lead to liver disease, such as cirrhosis or hepatocellular carcinoma (23, 27).

AAT is encoded by the SERPINA1 gene, located on chromosome 14q32.1 (15). Several allelic variations have been reported, with varying effects on AAT levels (48). These variants are annotated by their characteristic migration

patterns in isoelectic focusing analysis. Normal AAT migrates as a middle band (M), and other isoforms are classified as A-L for faster migrating bands and N-Z for slower migrating bands. A1AD is caused by one of several mutations of autosomal alleles that leads to low levels of wild-type Alpha-1-Antitrypsin circulation. The most clinically relevant of these mutations is the Z (Glu342Lys) mutation, which leads to drastically reduced levels of circulating AAT (20% of wildtype, in patients that are homozygous for the Z allele) (42). While the causative agent of this disease is understood as genetic mutations, there is little explanation for the wide variation of penetrance and disease severity seen in patients with this mutation (27, 42). As such, an increased understanding of the regulation of both the wild-type and disease form of AAT will be invaluable to the patient community.

Much attention has been given to genomic variation in the SERPINA1 gene and how this variation can account for disease phenotypes. There is, however, a paucity of information on post-trascriptional regulation of SERPINA1. The human transcriptome gives rise to multiple isoforms of a number of genes, and variation in pre-mRNA and mRNA transcripts can contribute significantly to variation gene expression. Here we focus on variation in the 3' Untranslated Region (UTR) of SERPINA1, the site of a significant portion of post-transcriptional regulation.

The average 3'UTR in the human transcriptome is roughly 1000 nucleotides (nt) in length (65). SERPINA1 has been reported to have an

extremely long (1700 nt) 3'UTR. This region of non-coding RNA can confer regulation in multiple ways, using both *cis* factors encoded in the UTR sequence as well as *trans* factors such as RNA-binding proteins to increase or decrease translation of the mRNA (20 and references therein, 50). 3'UTR-mediated regulation is necessarily sequence-specific, as it depends greatly on the presence and availability of binding sites for regulatory complexes. As such, truncations or elongations of a transcript's 3'UTR can significantly alter post-transcriptional regulation, either ablating or adding regulatory sites. Additionally, single-nucleotide changes in the 3'UTR can disrupt *cis* regulator factors such as signal sequences for processing machinery and nuclear export or binding sites of *trans* regulatory factors such as microRNAs (miRNAs) (21, 25, 54).

One of the most prevalent mechanisms of 3'UTR-dependent regulation is Alternative Poly-Adenylation (APA), which affects an estimated 70% of transcripts (reviewed in 20). With the exception of a few mRNAs, such as histone mRNAs, most mature eukaryotic mRNAs end in a non-templated poly-adenosine (poly-A) tract. This non-templated end is produced through the recognition a six nucleotide motif, typically AAUAAA (5), within the 3'UTR by two protein complexes upon the transcript's emergence from RNA polymerase II (RNA pol II) (5, 20). Recognition and binding to this sequence is followed by cleavage of the transcript roughly 40 nt downstream, and the addition of the poly-A tail by a poly-A-polymerase. APA causes transcripts with varying 3'UTR length and variation in poly-A tail length due to the recognition of differing signal sequences within the

same 3'UTR. There are multiple potential poly-A signal sequences in the SERPINA1 3'UTR (Figure 4B). The most distal of these is a canonical AAUAAA hexamer resulting in a 1700 UTR, but an AUUAAA sequence ~70nt downstream of the stop codon may also be used (20). Reports have varied on the identity of the predominant SERPINA1 3'UTR. The variation between multiple SERPINA1 UTRs could account for differential post-transcriptional regulation, but to our knowledge, no reports have investigated the impact of variation in SERPINA 3'UTRs.

Single nucleotide polymorphisms (SNPs) can also generate variation in 3'UTR transcripts. "Common" SNPs occur in 1% or more of the population, and current estimates report 10 million SNPs in the human genome. SNPs in coding regions of genes are often given much attention for their ability to significantly alter amino acid sequences. However, roughly 93% of SNPs currently known are in non-coding regions. Though non-coding SNPs do not affect protein identity by changing amino acid sequences, they can significantly alter gene expression through changes in transcriptional or post-transcriptional regulation. SNPs in the 5' UTR can alter promoter or enhancer sequences, greatly affecting transcription efficiency. SNPs in the 3'UTR can alter binding sites of miRNA or RNA-binding proteins and impact 3'UTR-dependent post-transcriptional regulation to great consequence. Indeed, multiple 3'UTR SNPs have been reported to disrupt or disease-associated miRNA binding sites and alter gene expression through increases or decreases in miRNA-mediated gene repression.

miRNAs are of specific interest to liver-health, as several miRNAs have been implicated in a wide range of liver processes and disease. To our knowledge, no miRNAs have been identified as yet as regulators of AAT. Recently, miR-199 was identified as a key regulator of the unfolded protein response to ER stress caused by the ZZ isoform of AAT (31, 32). This work contributes to a larger understanding post-transcriptional regulation surrounding SERPINA1. It does not, however, elucidate pathways of SERPINA1 regulation, focusing instead on downstream effects of misfolded SERPINA1 gene product. Our work demonstrates a significant step forward in the understanding of SerpinA1 post-transcriptional regulation, as it validates previous work confirming the SerpinA1 3'UTR, and is the first large-scale investigation of direct and indirect regulation of secreted AAT protein by miRNAs.

Results: Effects of SERPINA1 3'UTR and variants on AAT secretion in the liver

MiRNAs do not play a role in secretion of wild-type AAT

We first determined candidate miRNAs as possible regulators of secreted AAT. As AAT is highly expressed in the liver, and the liver is the site of most disease relevance, it stands to reason that abundant liver miRNAs would be prime candidates to test as regulators of secreted AAT. To find these, we used miRNA-seq data deposited in The Cancer Genome Atlas to find the most abundant miRNAs expressed in normal liver cells (Supplementary Table 1).

Scoring miRNAs based on number of reads detected, we calculated the top fifty candidates across multiple data sets.

To test whether or not these candidates regulate AAT, we first examined whether over-expression of these mimics can significantly alter secreted AAT levels in tissue culture. We transfected Huh-7 or HepG2 liver cell lines with either mimics for each of the top fifty liver miRNAs, negative control mimics, anti-SERPINA1 siRNA, or negative control siRNA obtained from Sigma Aldrich. Two days post-transfection, RNA and cell culture supernatant were harvested for testing. We assayed SERPINA1 transcript levels by qPCR to confirm efficient delivery of these small RNAs, and knockdown of SERPINA1 transcripts in the presence of anti-SERPINA1 siRNA was confirmed (Figure 1A). To confirm that decrease in SERPINA1 transcript did indeed cause a decrease in secreted protein in both cell lines, we compared AAT levels in both mock, negative control siRNA, and anti-SERPINA1 siRNA-transfected cells (Figure 1B). Upon confirming of a reduction of roughly 80% in secreted protein by Enzyme Linked Immunosorbance Assay (ELISA), we assayed AAT levels for all miRNA mimics transfections by ELISA. Secreted AAT levels were assayed by ELISA from cell culture supernatant to determine whether overexpression of any of the candidate affects levels of secreted AAT positively or negatively. All values were normalized to a negative control miRNA mimic and to NanoLuc, an irrelevant secreted protein luciferase reporter, to control for non-specific decreases in secreted protein.

As an initial screen, we tested all fifty miRNA mimics, and any candidates that scored 20% higher or lower than the negative control mimic were selected for further screening (Figure 1C). Six candidates, Let-7a, miR-7a, miR-23, miR-26, miR-27, and miR-182, did score by these criteria in our initial screen, and were subsequently assayed by qPCR for changes in transcript levels as well as in secreted protein (Figures 2A and 2B respectively). None of these six candidates screened consistently altered levels of either SERPINA1 transcripts or secreted AAT levels in follow-up screens, suggesting that overexpression of these miRNAs do not detectably alter AAT levels.

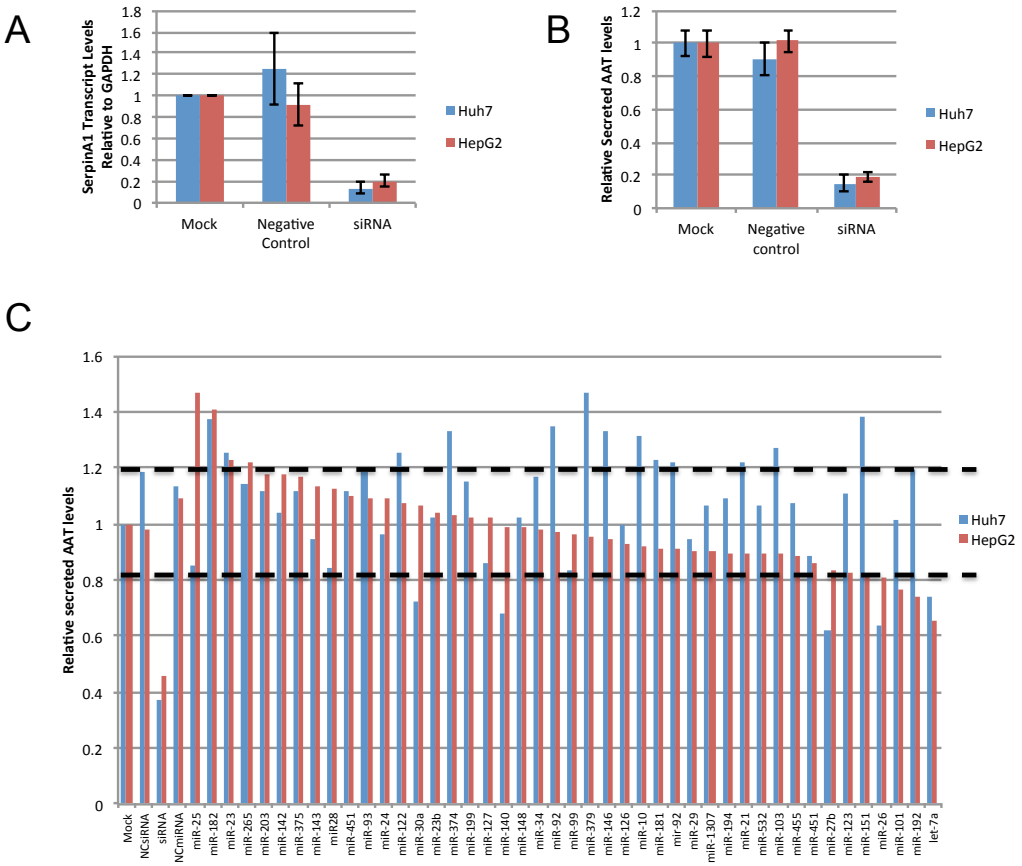


Figure 1.1 Effects of noncoding RNAs on SerpinA1 transcript and gene product levels
(A) Quantitative real-time PCR analysis of SERPINA1 transcript levels. Two liver cell lines (Huh-7 and HepG2) were transfected with mock transfection conditions (RNAiMax and water), negative control siRNA, or anti-SERPINA1 siRNA, and SERPINA1 transcript levels were assayed by SYBR Green qPCR. Transcript levels were normalized to GAPDH transcript levels. Bars represent the mean SERPINA1 +/- SD (n=3) from three experiments in which transfections were performed in triplicate. (B) ELISA analysis of secreted AAT protein. Two liver cell lines (Huh-7 and HepG2) were transfected with mock transfection conditions (RNAiMax and water), negative control siRNA, or anti-SERPINA1 siRNA. Forty-eight hours post transfection, cell culture supernatant was harvested and secreted AAT levels were assayed by Abcam ELISA. Secreted AAT levels normalized to mock transfection. Bars represent the mean relative AAT +/- SD (n=3) from three experiments in which transfections were performed in triplicate. (C) ELISA analysis of secreted AAT protein in miRNA mimic-transfected cells. Two liver cell lines (Huh-7 and HepG2) were transfected with mock transfection conditions (RNAiMax and water), negative control siRNA, anti-SERPINA1 siRNA, negative control miRNA mimic, or one of fifty liver miRNA mimics. Forty-eight hours post transfection, cell culture supernatant was harvested and secreted AAT levels were assayed by ELISA (Abcam ab108799). Secreted AAT levels normalized to mock transfection Bars represent relative AAT levels from one experiment in which single transfections were performed per liver miRNA.

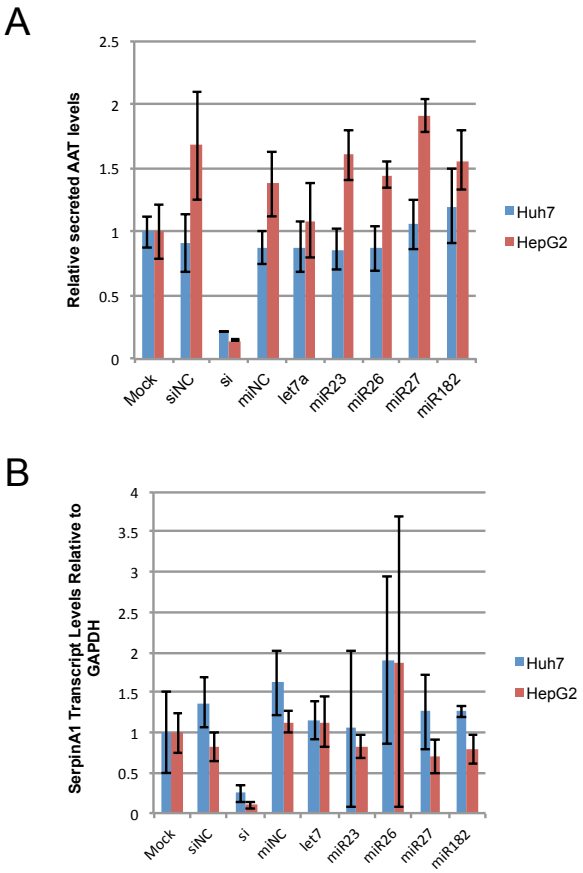


Figure 1.2. Secondary screen of initial miRNA hits does not reveal direct or indirect AAT regulators.
(A) ELISA analysis of secreted AAT protein. Two liver cell lines (Huh-7 and HepG2) were transfected with mock transfection conditions (RNAiMax and water), negative control siRNA, anti-SERPINA1 siRNA, negative control miRNA mimic, or one of six "hit" liver miRNA mimics. Forty-eight hours post transfection, cell culture supernatant was harvested and secreted AAT levels were assayed by Abcam ELISA. Secreted AAT levels normalized to mock transfection. Bars for HepG2 cells represent the mean relative AAT +/- SD (n=2) from two experiments in which transfections were performed in triplicate. Bars for Huh7 cells represent the relative AAT +/- SD (n=1) from one experiment in which transfections were performed in triplicate. (B) Quantitative real-time PCR analysis of SERPINA1 transcript levels. Two liver cell lines (Huh-7 and HepG2) were transfected with mock transfection conditions (RNAiMax and water), negative control siRNA, anti-SERPINA1 siRNA, negative control miRNA mimic, or one of six "hit" liver miRNA mimics. SERPINA1 transcript levels were assayed by SYBR Green qPCR. . Bars for HepG2 cells represent the mean SERPINA1 +/- SD (n=2) from two experiments in which transfections were performed in triplicate. Bars for Huh7 cells represent SERPINA1 +/- SD from one experiment in which transfections were performed in triplicate.

Next, to test whether any liver miRNAs were responsible for regulation of secreted protein, regardless of expression level, we infected Huh-7 liver cells with either an adenoviral vector expressing GFP (Ad-GFP) or the same vector with a fused poxvirus protein (Ad-GFP-VP55) which globally degrades miRNAs (2, 3). To confirm that miRNA expression was significantly reduced, RNA from infected cells was harvested and assayed for miRNA expression by Northern blot analysis. MiRNA expression was reduced in Ad-GFP-VP55-infected cells (Figure 3A) by roughly 85%. Cell supernatant from the same infected cells were assayed by ELISA to determine relative secreted AAT levels, alongside cells transfected with either negative control siRNA or anti-SERPINA1 siRNA (Figure 3B). No difference in secreted AAT levels was observed between control-infected or Ad-GFP-VP55-infected cells, suggesting that secretion of wild-type AAT is not affected by under-expression of miRNAs. Results of both over- and under-expression analysis suggest that in the cell lines examined, miRNAs do not play a role in secretion of wild-type AAT.

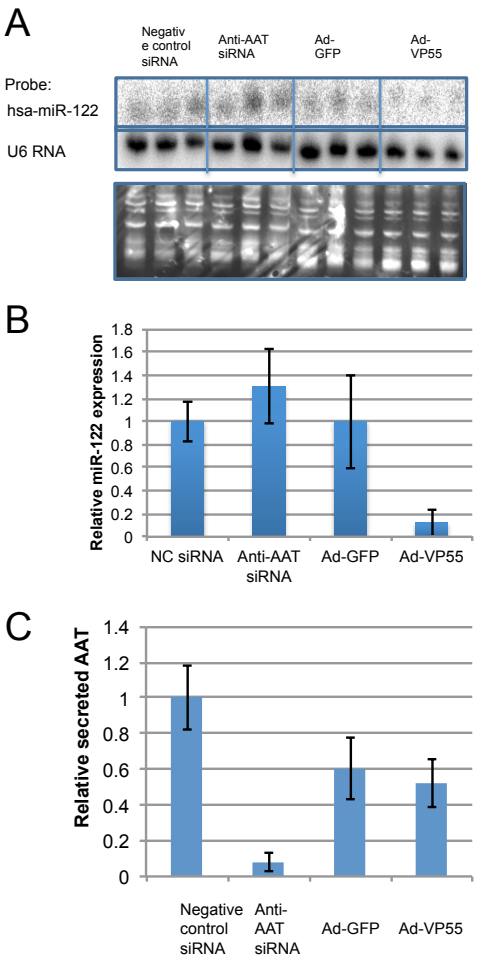


Figure 1.3. Reduced miRNA expression does not significantly affect secreted AAT levels. (A) Northern blot analysis showing decrease in miRNA expression in AdGFPVP55-infected cells. Huh-7 cells were infected (MOI of 500) with either control adenoviral vector AdGFP or a VP55 vector AdGFPVP55, or transfected with either negative control siRNA (Sigma-Aldrich) or SERPINA1 siRNA. Forty-eight hours post-infection or post-transfection, total RNA was extracted and Northern blot analysis probing for liver miRNA miR-122 was performed. (B)Quantitative analysis of reduction in signal from Northern blot in Figure 3.A. Density of miR-122 was quantified for each lane and relative to signal for U6 RNA. Averages per treatment are shown relative to negative controls (n=3) Bars represent standard deviation from mean.(C) ELISA analysis of secreted protein from indicated cells. As indicated above, . Huh-7 cells were infected (MOI of 500) with either control adenoviral vector AdGFP or a VP55 vector AdGFPVP55, or transfected with either negative control siRNA (Sigma-Aldrich) or SERPINA1 siRNA, and cell culture supernatant was assayed for secreted protein levels 48 hours post infection or post transfection. Bars represent the mean relative AAT +/- SD (n=2) from two experiments in which transfections were performed in triplicate.

Regulation of gene expression is independent of the SERPINA1 3'UTR

To investigate the effects of the SERPINA1 3'UTR on protein expression, luciferase reporters were made with either the parental pSICHECK 2 plasmid's 3'UTR or the SERPINA1 3'UTR cloned behind a Renilla luciferase gene. Liver and non-liver cells were transfected with these reporter plasmids and Renilla luciferase expression relative to control firefly luciferase was assayed to determine liver-specific regulation of the SERPINA1 3'UTR (Figure 4A). However, no difference in expression was observed between liver and non-liver cells, suggesting that no liver-specific trans factors are responsible for regulation of the SERPINA1 transcript in the cell lines tested.

Recently published data (58, 75) suggests that, contrary to previous reports, the predominant 3'UTR of SERPINA1 is much shorter than previously believed (Figure 4B). A non-canonical poly-adenylation signal sequence roughly 80 nucleotides from the stop codon may be read by poly-A machinery to produce a 3'UTR that is less than a tenth of the 1.7 Kb UTR. Recently, RNA-seq data deposited in ENSEMBLE (58) showed that the predominant form in liver tissue is indeed this shorter 3'UTR. To determine whether the longer transcript, which is still produced in low amounts, could be regulated we made two poly-A mutant reporters: 1) a two nucleotide mutation in the upstream signal sequence and 2) a deletion of the upstream signal sequence. We confirmed that these mutations make the full-length 3'UTR in greater abundance than the wild-type reporter using 3'RACE primers for nested PCR. Indeed, relative to the wild-type reporter,

both mutant reporters do favor the full-length SERPINA1 3'UTR (Figure 4C). While we have not confirmed that the short 3'UTR is produced in decreased abundance relative to the wild-type reporters, our initial PCR results give us confidence that the APA reporters are working as expected. However, both of these mutant reporters appear to repress expression to the same degree as the wildtype reporter (Figure 4D). This suggests that SERPINA1 alternative poly-adenylation has no regulatory role in the systems we analyzed.

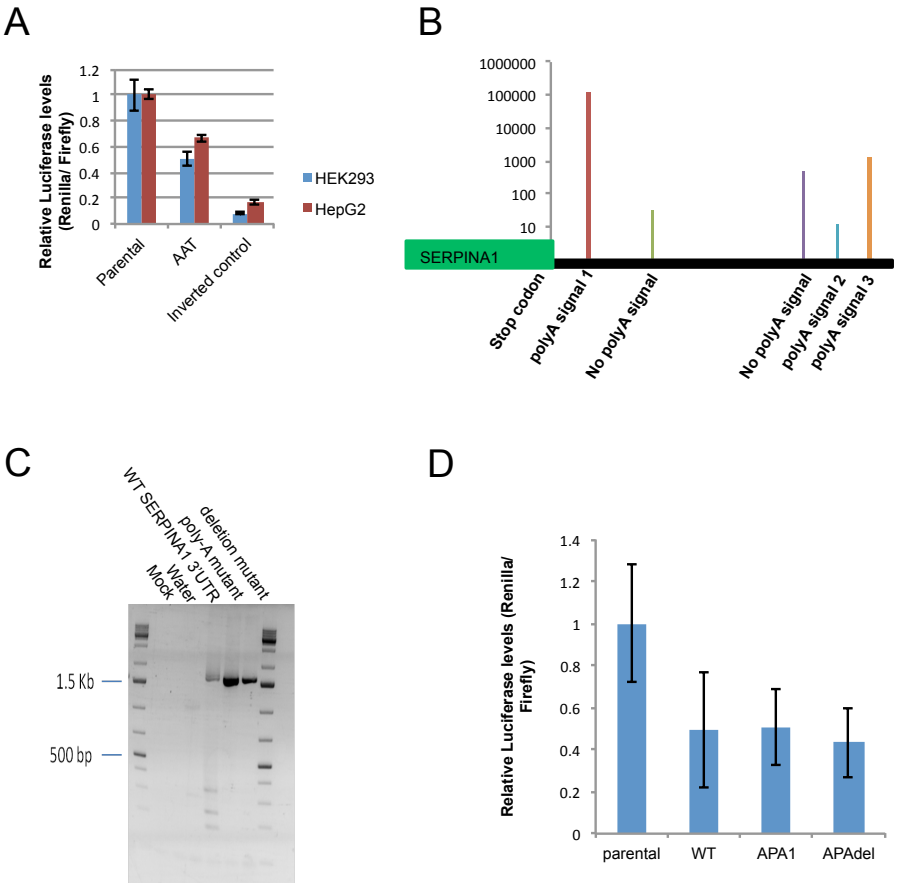


Figure 1.4. AAT 3'UTR does not confer regulation in a liver or miRNA-specific context.

(A) Luciferase assay measuring 3'UTR-dependent contribution to gene expression in liver and non-liver cells. Parental dual luciferase vector or modified vector with either AAT 3'UTR or inverted AAT 3'UTR (length control) cloned downstream of Renilla luciferase coding sequence were transfected into HEK293 or HepG2 cells. Renilla luciferase levels were assayed for gene expression with firefly luciferase expression serving as a transfection control. Bars represent the mean ratio (Renilla/Firefly) +/- SD (n=3) from three experiments in which transfections were performed in triplicate. (B) PolyA signal usage from liver tissue: graphical representation of SERPINA13'UTR length using data from APASdb. Bars represent reads in log scale mapping to regions with x axis representing distance from stop codon. (C) Nested PCR detection of Renilla luciferase reporter transcripts: HEK293T cells were transfected with indicated reporters. 48 hours post-transfection, ttal RNA was harvested, Dnase I treated to remove plasmid contamination, and reverse-transcribed using anchored oligo-d(T) primers. SERPINA1 3'UTR was amplified using primers for full-length 3'UTR. (D) Luciferase assay measuring contribution of Alternative PolyAdenylation to 3'UTR-mediated regulation of gene expression. Parental dual luciferase vector or modified vector with either AAT 3'UTR, polyAdenylation sequence mutant, or poly-adenylation deletion mutant AAT 3'UTR cloned downstream of Renilla luciferase coding sequence were transfected into HEK293 cells. Renilla luciferase levels were assayed for gene expression with firefly luciferase expression serving as a transfection control. Bars represent the mean ratio (Renilla/Firefly) +/- SD (n=3) from three experiments in which transfections were performed in triplicate.

The SERPINA1 3'UTR of ZZ-related A1AD patients do not contain common SNPs

In order to determine whether or not SNPs present in the 3'UTR of SERPINA1 could modulate disease severity, we analyzed sequence of the SERPINA1 3'UTR region of a large number of A1AD patients. We obtained genomic DNA samples from 250 patients with varying degrees of A1AD, with associated clinical data, from the Alpha-1-Antitrypson Foundation DNA Bank. We used Sanger sequencing to sequence the wild-type SERPINA1 3'UTR of each of these patients. The SERPINA1 3'UTR was first amplified out of genomic DNA by polymerase chain reaction, then submitted for sequencing through the University of Texas at Austin DNA Sequencing Facility. To eliminate the possibility of PCR-contamination, these samples were amplified on different days, with no template control reactions to ensure that the experimenter's DNA was not being amplified. An assay was developed to identify the sex of the patient by amplifying a Y-specific gene, and used to corroborate the unique identity of several patients (Figure 5A) as a further control. Additionally, an alternate genomic region from a different chromosome was sequenced to verify the capability of our approach to detect SNPs in diverse DNA populations. As expected, patients did have unique SNPs in common SNP regions (Figure 5B), but no unique SNPs were found in the genomic region of the SERPINA1 3'UTR of any of the patients (Table 1). This is likely due to the relatively recent occurrence of the ZZ allele responsible for A1AD seen in all sequenced patients.

SNP:	Patient population	Control DNA	Expected Frequency
SERPINA1 3'UTR	0	-	N/A
AAT-linked locus	100%	25%	25%
Unlinked SNP	~50%	50%	50%

Table 1.1: Sequencing of A1AD patients 3'UTR
Patient and non-patient populations sequenced for disease-linked SNPs as well as SNPs on unrelated chromosomes. Patients were identical in gene regions proximal to A1AD-causative mutation, and diverse in distal regions.

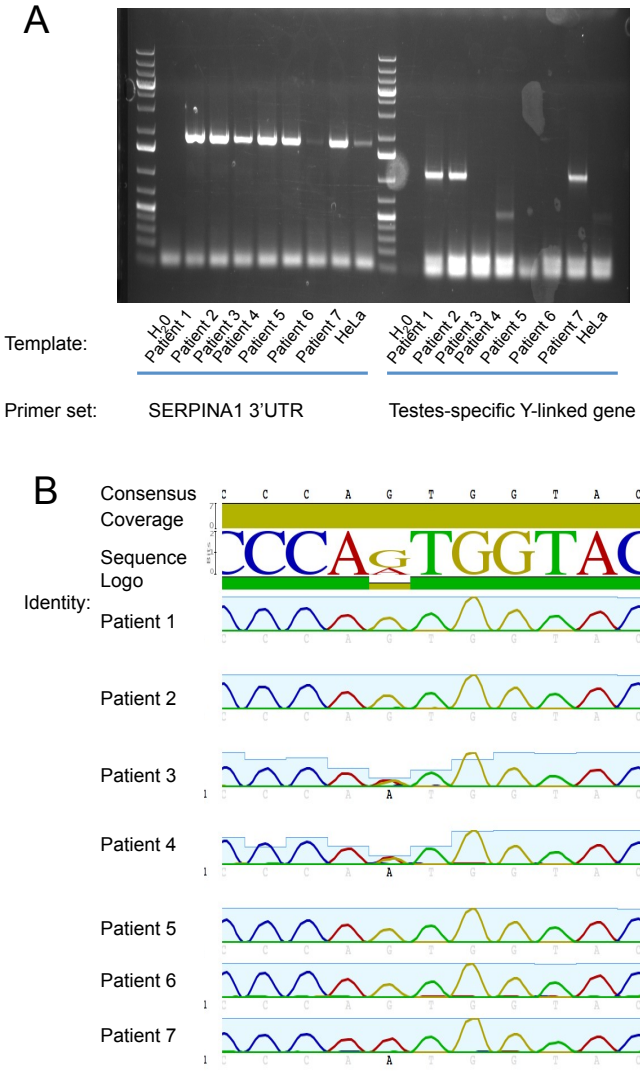


Figure 1.5. Sequencing of A1AD patient genome reveals no link between 3'UTR SNPs and disease severity.

(A) Verification of unique identity of sequencing samples through amplification of Y-linked gene. PCR amplification of both AAT 3'UTR and Testes-specific Y-linked gene 1 (TSYP1) visualized by ethidium staining from genomic DNA. All input templates yield a band using AAT 3'UTR primer pair, but only male patients show characteristic 492 nt-length banding with Y-chromosome specific primer pair. Water and genomic DNA from HeLa (female) cells are provided as negative controls. (B) Sequencing chromatograms of SNP region on non-SERPINA1-related region on chromosome 12. Sequence logo size related to relative frequency of nucleotide in pool of patient sequences.

Discussion

Alpha-1 antitrypsin deficiency disorder is a disease that can cause a wide range of health defects, from asthma to liver failure. Disease pathologies are largely caused by aberrant protein folding due to genetic mutations in the coding sequence of the SERPINA1 gene. We hypothesized that some of this variation could be explained by variation in 3'UTR-mediated post-transcriptional regulation.

Variations in 3'UTR-mediated post-transcriptional regulation can be caused by multiple *cis* and *trans* factors involved in gene expression. As the liver is the primary source of circulating AAT, we first tested whether or not liver miRNAs are *trans* regulators of SERPINA1 gene expression. We determined the top 50 liver-relevant miRNAs from over 1000 total human miRNAs using large datasets of small-RNA sequencing of healthy liver tissues from The Cancer Genome Atlas. We tested whether or not overexpression of these miRNAs could contribute directly or indirectly to production of secreted AAT in cell culture models, scored by a 20% increase or decrease in secreted protein relative to mock transfected cells in two different liver cell lines. Three miRNAs, miR-25, miR-182, and let-7a scored according to our original criterion. Due to this unexpectedly low number, we relaxed our original screening threshold to include miR-23 and miR-26, which were just outside the original threshold. However, in subsequent screens with multiple replicates, each of these original hits failed to consistently affect secreted AAT levels in a consistent manner. Additionally, they

did not seem to greatly impact SERPINA1 transcript levels. These data suggest that, at least in cell culture models, miRNA overexpression does not significantly increase or decrease AAT levels. Next, we investigated whether decreases in overall miRNA expression could impact secreted AAT protein. Utilizing a novel strategy developed by the tenOever lab, we delivered a poxvirus protein that degrades host miRNAs nonspecifically levels adenoviral vector and analyzed subsequent changes in AAT relative to control infection. Again, we did not observe any significant changes in secreted protein level, suggesting that in our given timeframe, the effects of decreased miRNA activity are not sufficient to impact secreted AAT levels. These observations are consistent with recently published data that suggest the predominant SERPINA1 3'UTR is the short form. With roughly 100 nucleotides of sequence space, there is little room for miRNA docking sites. Importantly, these data do not reflect effects on misfolded protein, measuring instead only wildtype AAT levels. It is possible that miRNAs impact levels of mutant AAT. Indeed, some reports have implicated miR-199 as a regulator of the unfolded protein response to the ZZ isoform of AAT, though no direct regulators of misfolded AAT have yet been uncovered. Our data combined suggest that in a cell culture model, neither overexpression nor underexpression of miRNAs is sufficient to alter secreted protein levels.

Next we investigated whether the AAT 3'UTR, or variant transcripts, could contribute to regulation of gene expression. First we tested whether or not the AAT 3'UTR ould confer regulation in a liver-specific context. Using luciferase-

based reporters, we compared the AAT 3'UTR to a parental vector 3'UTR, or an inversed AAT 3'UTR as a length control. Consistent with reports of a shorter 3'UTR, we observed some amount of repression of Renilla luciferase gene product with the AAT 3'UTR relative to a parental vector 3'UTR of around 40-50%. This repression is far less than that observed with the inverted 3'UTR size control. Additionally, we did not observe a significant difference in 3'UTR-dependent repression between liver and non-liver cell lines. These data suggest there are no liver-specific trans factors that affect 3'UTR-mediated repression in cell culture models. Next, we tested whether or not alternative transcripts could confer additional regulation. As alternative poly-adenlyation is one of the most biologically significant sources of 3'UTR variation, we created poly-adenylation mutant reporters and verified that full-length AAT 3'UTRs are indeed produced in increasing abundance relative to our wild-type reporter. Surprisingly, the full-length (1.7Kb) AAT 3'UTR does not confer additional repression relative to the more common, short 3'UTR when transfected into cells and assayed for Renilla luciferase levels. These reporters were tested in both liver and non-liver cells (data not shown), suggesting the possibility that the 3'UTR of SERPINA1, a major gene product of the liver, is not regulated by either liver-specific *trans* factor or *cis* factors leading to alternative polyadenylation.

Finally, to test whether or not genetic variation in the AAT 3'UTR could contribute to disease severity, we sequenced the 3'UTR of 250 A1AD patients that are homozygous for the Z allele. We grouped these patients by disease

severity ranked on two measures: levels of inflammatory response protein CRP, and forced expiratory volume (FEV), a measure of respiratory distress. To our surprise, all 250 patient genomes have identical AAT 3'UTR regions. We additionally sequenced several patient genomes downstream of the wildtype AAT 3'UTR, at a locus at which 25% of the general population has a single nucleotide polymorphism (SNP) according to NCBI's dbSNP. In contrast to the general population, 100% of the patient UTRs sampled contained the same allele. This is likely due to the proximity of this allele locus to the shared ZZ allele in all tested A1AD patients. The ZZ allele arose from a single founder in a Viking colony an estimated twenty generations ago, and genetic linkage to any proximal SNPs is unlikely to be broken in such a short time (11). Importantly, a common SNP reported in 50% of the general population on an alternate (unlinked) chromosome, rs4149057, was equally represented in the patient population, demonstrating that unique SNPs are detectable in said population. These data suggest that SERPINA1 3'UTR variation does not contribute to A1AD disease severity.

In conclusion, we have found no link between overexpression over liver-expressed miRNAs and wildtype AAT regulation. Additionally, we found underexpression of all miRNAs does not impact wildtype AAT secretion. We found that the predominant SERPINA1 3'UTR is indeed atypically short, and does not confer gene regulation in a liver-specific manner. Finally, we determined

SERPINA1 3’UTR variation does not contribute to A1AD severity, likely due to genetic linkage to the causative ZZ allele.

Chapter 3: Identification of virus-encoded miRNAs in divergent Papillomaviruses

Introduction: *Viral miRNAs and Papillomaviruses*

Papillomaviruses (PVs) comprise a large family of circular double-stranded DNA viruses. Numerous PV genomes have been described including over 200 human PV (HPV) types. A minority of these are known as carcinogenic agents (80, 72) however only a small fraction of hosts infected with these high risk types will go on to develop high grade lesions. It remains incompletely understood what factors dictate whether or not HPV infection will develop into malignant cancer. Further, it is unclear why HPVs that share a high level of sequence similarity can have stark differences in tropism and infect different regions of the body. Developing a better understanding of PV gene products and regulation of their expression within the larger context of the PV family provides a foundation for deciphering the mechanisms controlling the differential outcomes of infection.

An emerging class of viral gene products that are found in select virus families are small regulatory RNAs called microRNAs (miRNAs). Over 300 viral encoded miRNAs have been described, all from viruses able to undergo long term persistent infection (36, 63, 70, and references therein). Most of these viruses have DNA genomes including the herpes, polyoma, and anello virus families. However, delta and foamy retroviruses also encode miRNAs (9, 37, 70). One likely role of viral miRNAs is to foster productive long-term interactions

within the host (reviewed in 70). In this regard, at least some members of the PV family would be expected to encode miRNAs. However, to date, no credible examples of PV canonical miRNAs have been described. Two studies examining fully infectious experimental systems of the high-risk HPV types do not find evidence for virus-encoded miRNAs (78). Further, a study of latent and productive HPV31 cycles of replication concluded that this high-risk HPV does not encode miRNAs (12). There have been reports in transformed cells of small RNAs from high risk PVs such as HPVs16 &18, however these studies did not demonstrate a connection of PV-derived small RNAs to the miRNA biogenesis (Dicer/Drosha) or effector (RISC) machineries (57, 71). Therefore, these RNAs likely represent degradation fragments derived from the turnover of longer PV transcripts. In contrast, it is well documented that PV infection and individual PV gene products can alter the host miRNA repertoire, likely contributing to the biology of cancer (26, 30, 52, 73, 78). Furthermore, at least one PV, HPV31, utilizes a host miRNA to directly regulate early viral gene expression. Thus, what emerges is that although host miRNAs are involved in the PV life cycle and pathogenesis, of the few PV types that have been examined, no canonical PV miRNAs are yet established.

One barrier to discovery of PV miRNAs is the dearth of facile fully-infectious laboratory systems. There are experimental systems established for a few PV types (approximately <5) (18), but technological barriers have limited any comprehensive large-scale study of viral miRNAs in the majority of PV types.

Here we describe a new wet bench approach for the discovery of miRNAs that assays numerous viruses in parallel for the ability to express miRNAs. We identify bona fide PV miRNAs encoded by divergent PVs, and clearly demonstrate these miRNAs depend on canonical miRNA biogenesis effector machinery. Our further analysis rules out abundant canonical viral miRNAs in cancers associated with high-risk PV (HPVs 16, 18, and 31) infection. These findings resolve the issue of PV miRNAs and further the notion of miRNA importance to persistent virus infection.

RESULTS

Proof of concept for miDGE on a herpesviral genome.

To identify miRNA genes in situations where transcripts are not easily obtainable, we developed the approach of miRNA Discovery by forced Genomic Expression (miDGE). miDGE relies on generating a library of numerous overlapping genomic segments of DNA from a particular organism or locus and subcloning them behind a heterologous RNA polymerase (RNP) II promoter (Figure 1). The concept relies on the principle that miRNA genes are compact and should be readily expressed by heterologous upstream RNP II promoters, or in the rare cases that a primary miRNA transcript is driven by RNP III, that these promoters are small and proximal to the miRNA gene so as to be included in miDGE library constructs. The miDGE library is then transfected into mammalian cells and small RNA is harvested and sequenced. Next, we apply bio-

computational methods to identify miRNA candidates whose transcripts display the hallmarks of processing by the miRNA biogenesis machinery. Finally, candidate miRNAs are validated via a series of molecular assays to establish biogenesis via the canonical miRNA processing machinery and activity within RISC, the miRNA silencing machinery.

To test the effectiveness of miDGE approach, we first focused on a single larger genome virus, the herpesvirus Japanese Macaque Rhadinovirus (JMRV). JMRV is a gamma-2 herpesvirus with genomic sequence similar to the highly related Rhesus rhadino virus (RRV). When we initiated these studies, it was not yet known if JMRV encoded miRNAs, although it would be expected to given the numerous precursor miRNAs (pre-miRNAs) identified in RRV that share high sequence similarity with JMRV (64). Indeed, work from Skalsky et al. has now identified 15 novel viral miRNA encoded by JMRV. Thus, JMRV serves as proof-of-principle test genome to evaluate miDGE. miDGE analysis correctly identified all fifteen known miRNAs in JMRV (data not shown). Importantly, there were no other high-confidence candidate miRNAs called within this 20 kB region. Thus, miDGE has a low background rate, at least in the context of a herpesviral genome and can successfully identify bona fide miRNAs.

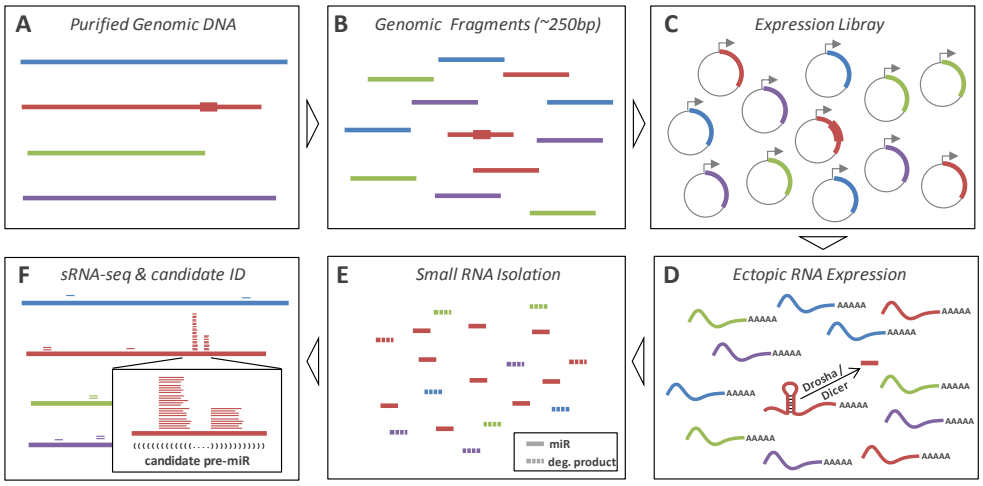


Figure 2.1: Overview of the miDGE (miRNA Discovery by forced Genomic Expression) Methodology

(A) Pools of purified viral genomes are subjected to fragmentation via sonication or limited digest (average target size of fragments approx. 250 bp) (**B**), follwed by cloning of subgenomic fragments to create an expression library (**C**). Thickened regions in A-C symbolize an unidentified pre-miRNA coding region present in one of the viral genomes. (**D**) After transient transfection, cloned subgenomic fragments are forcibly transcribed, allowing processing of pre-miRNA hairpins to produce mature miRNAs. Small RNAs purified from transfected cells (**E**) will contain mature miRNAs as well as random degradation products produced from library transcripts. To identify authentic miRNAs, small RNAs are sequenced and mapped back to viral genome pools (**F**). The majority of degradation products exhibits a random distribution, whereas miRNA products produce distinct pileups at pre-miRNA loci. Secondary structure prediction can then be used to identify characteristic pre-miRNA hairpin structures (indicated by bracket notation in F) at such loci. In this study, we used the miRDeep2 package to identify miRNA candidates.

Identification of candidate PV miRNAs.

Our goal was to screen numerous PV genomes representing diverse clades in the PV family. To accomplish this, we collected 125 cloned PV genomes from both human and non-human animal sources. Our collection contained representatives from the majority of known PV clades (Figure 2A). We then grouped subsets of these plasmids containing the various PV genomes into libraries. For four of the five libraries, we utilized 4-base pair cutter restriction enzymes to fractionate the PV DNA. For the fifth library, which contained mostly genomes also contained within the other libraries, we used sonication to fractionate the PV DNA. High throughput pyrosequencing of these DNA libraries revealed that we had good coverage of numerous genomes. Sixty one percent of genomes had greater than 95% coverage. We next conducted high- throughput pyrosequencing of small RNAs from cell transfected with these five libraries. We observed good coverage, with roughly 60% of the PV genomes having greater than 100% coverage at the RNA level and as expected, these correspond to the genomes best represented in our DNA miRNA expression libraries. These results demonstrate the plausibility of using our library methods to cover simultaneously a large number of PV genomes at the DNA and RNA level.

We next applied bio-computational methods to identify small RNA reads consistent with bona fide miRNAs. Our algorithmic approach prioritized those small RNAs of appropriate size (between 18-24 nucleotides) that had a read distribution as being plausibly derived from a pre-miRNA hairpin. This approach

identifies read density coverage showing two "plateaus" of read density coverage, where each plateau represents a miRNA derivative from either the 5' or 3' arm of a pre-miRNA hairpin (5p or 3p miRNA, respectively) (Figure2B). In between each plateau is a trough in density coverage where the terminal loop portion of the pre-miRNA is under-represented in our libraries due to terminal loops of pre-miRNAs not being stabilized in RISC as miRNA derivatives are. miDGE readily called miRNAs from the positive control polyomavirus genomes included in the libraries (Simian Virus (SV40) and Merkel cell polyomavirus (MPyV)). For the vast majority of the greater than 539,000 combined nucleotide PV genomic space covered, few miRNA candidates were called, consistent with a low false positive rate for miDGE. Importantly, no miRNAs were called from any of the plasmid sequences that did not map to cloned viral genomic regions, further emphasizing the low false positive rate. However, miDGE did call five high-scoring miRNA candidates deriving from four different PV genomes, HPV17, HPV37, HPV41, and FcPV, the last of which infects the common chaffinch *Fringilla coelebs* (Figure 2B-D). These results demonstrate the plausibility of using miDGE to identify miRNAs from complex expression libraries and suggest that at least some PVs may encode miRNAs.

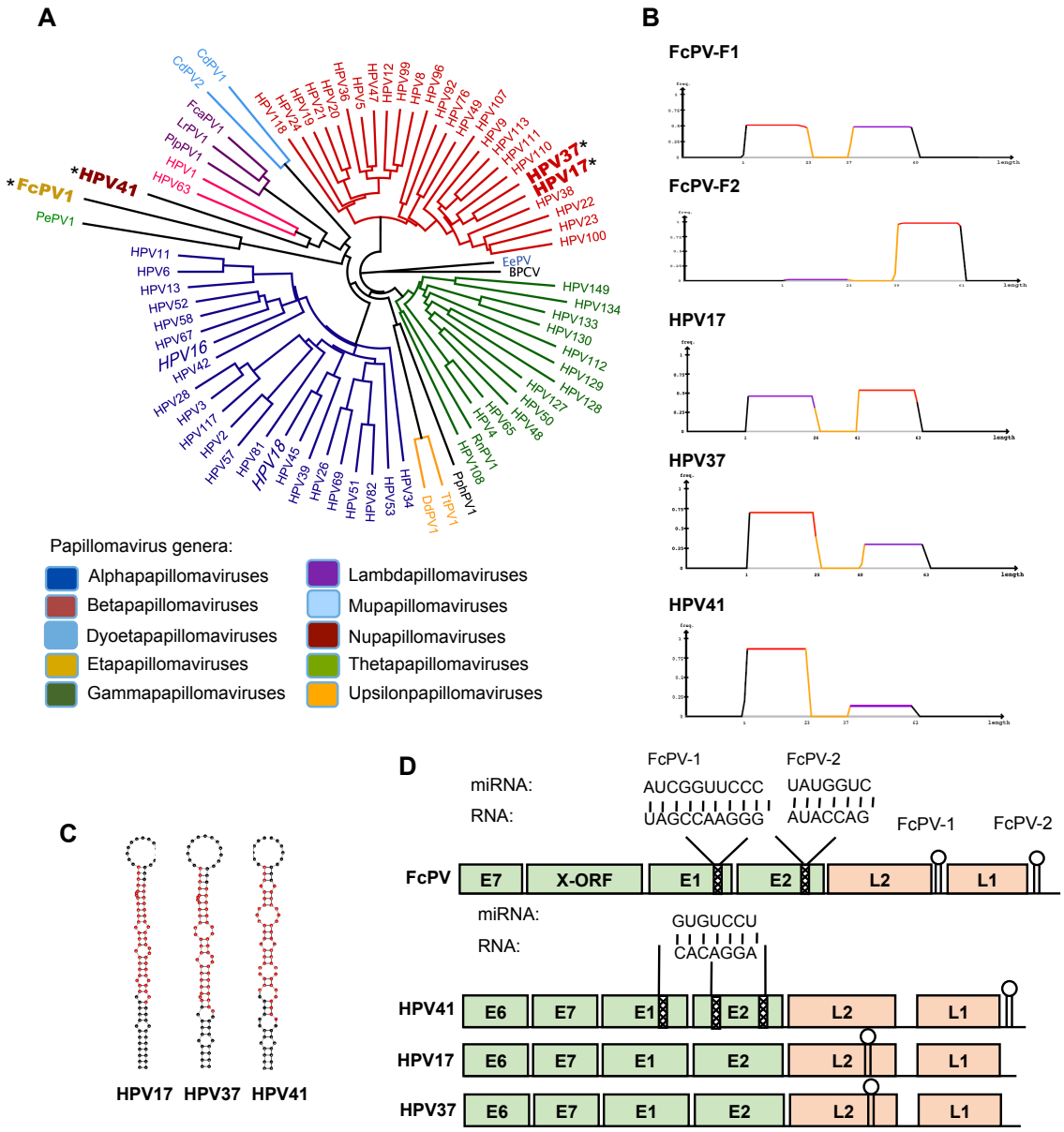


Figure 2.2: miDGE Assay Finds Five PV-encoded miRNA candidates.

A) Neighbor-joining tree calculated on alignment L1 nucleotide sequence of papillomaviruses included in miDGE library. Color indicates genus membership, with miRNA-encoding papillomaviruses in bold and cancer-associated papillomaviruses in italics. B) Frequency distrubition of the three miRNA candidates that were generated by MiRDeep2. C) Structures of HPV pre-miRNAs were predicted by minimal free energy folding using the RNAfold algorithm. The position of mature miRNAs is indicated by the red residues. D) The position of identified viral miRNAs is denoted by the hairpin shape. The identified seed sequence matches are noted at their respective positions with the sequences of the miRNA and potential targets.

Validation of PV miRNAs

The burden of proof for establishing bona fide miRNAs includes evidence of specific processing by the miRNA machinery and silencing activity within RISC. To vet the five PV candidate miRNAs, we first conducted northern blot analysis. Northern blot analysis can provide information about the size and processing of pre-miRNAs and derivative miRNAs. We cloned the candidate miRNA genes and flanking regions downstream of an RNP II promoter and transfected these plasmids into HEK293T cells. We harvested total RNA and conducted northern blot analysis. For each candidate, bands migrating at the appropriate size of typical of miRNAs were observed. Additionally, on most blots, a clear band consistent with a pre-miRNA was also observable. Thus, this analysis showed that all five candidates gave rise to banding patterns consistent with canonically processed miRNAs (Figure 3A).

To further validate these candidates, we investigated their biogenesis, activity and expression. To determine if the biogenesis of these candidate miRNAs required canonical miRNA machinery, we transfected our candidate miRNA constructs into cells with Drosha knocked down or that had Dicer knocked out (Figures 7 and 8). As expected, this analysis showed that positive control SV40 miRNAs were dependent on both Drosha and Dicer (Figure 3B and 3C, respectively). As a further control, a non-canonical Drosha-independent BLV miRNA (was not affected by knockdown of Drosha. All five candidate PV miRNAs showed reduced expression upon knockdown of Drosha (Figure 3B) and reduced

ratios of miRNA:pre-miRNA in the absence of Dicer (Fig 3C). These results conclusively demonstrate that the five candidate PV miRNAs derive from canonical miRNA biogenesis. We next generated luciferase-based RISC reporters for each viral miRNA candidate with two perfectly complementary sequences to determine if these candidate miRNAs are active in RISC. Transfection of these reporters with the candidate miRNA expression vectors resulted in specific knockdown of luciferase in each of the five reporters when co-transfected with respective miRNA expression vectors, but not in negative control reporters that had three nucleotide substitutions in the miRNA docking sites (Fig. 4A). These results suggest that each of the identified miRNAs are active in RISC. Finally, because FcPV infection gives rise to easily identifiable dense hyperplastic lesions on the feet of chaffinches, we were able to obtain RNA from host tissue infected by FcPV. Using qRT-PCR, multiple replicate experiments showed that two out of two diseased birds had detectable levels of at least one FcPV viral miRNA in FcPV-associated lesions. However, FcPV miRNAs were not detected in negative control tissues from the chests of these birds. We also obtained limited RNA from a third diseased bird from which we were only able to perform a single replicate for FcPV1, which gave consistent results (Table 1). Combined, the above results demonstrate that these candidates are bona fide PV miRNAs that derive from canonical biogenesis, are active in RISC and at least some are detectable *in vivo* in PV-associated disease tissue. In total, these results demonstrate that of our candidates, the FcPV miRNAs can indeed be

classified as bona fide miRNA, as they derive from canonical biogenesis, are active in RISC and are detectable *in vivo* in PV-associated disease tissue. Additionally, of the HPV miRNA candidates, we can say with high confidence they are likely viral miRNAs, meeting all criteria with the exception of being tested and established as being present in infected cells. In keeping with the miRNA naming convention set forth by miRBase, for the remainder of this paper, we name these miRNAs: FcPV miR-F1, FcPV miR-F2, HPV17 miR-H1, HPV37 miR-H1 and HPV41 miR-H1. We have deposited the FcPV miR-F1, FcPV miR-F2 pre-miRNA/miRNA sequences in miRBase and we note that until HPV miRNAs are confirmed in infected cells, these sequences will be withheld from miRBase.

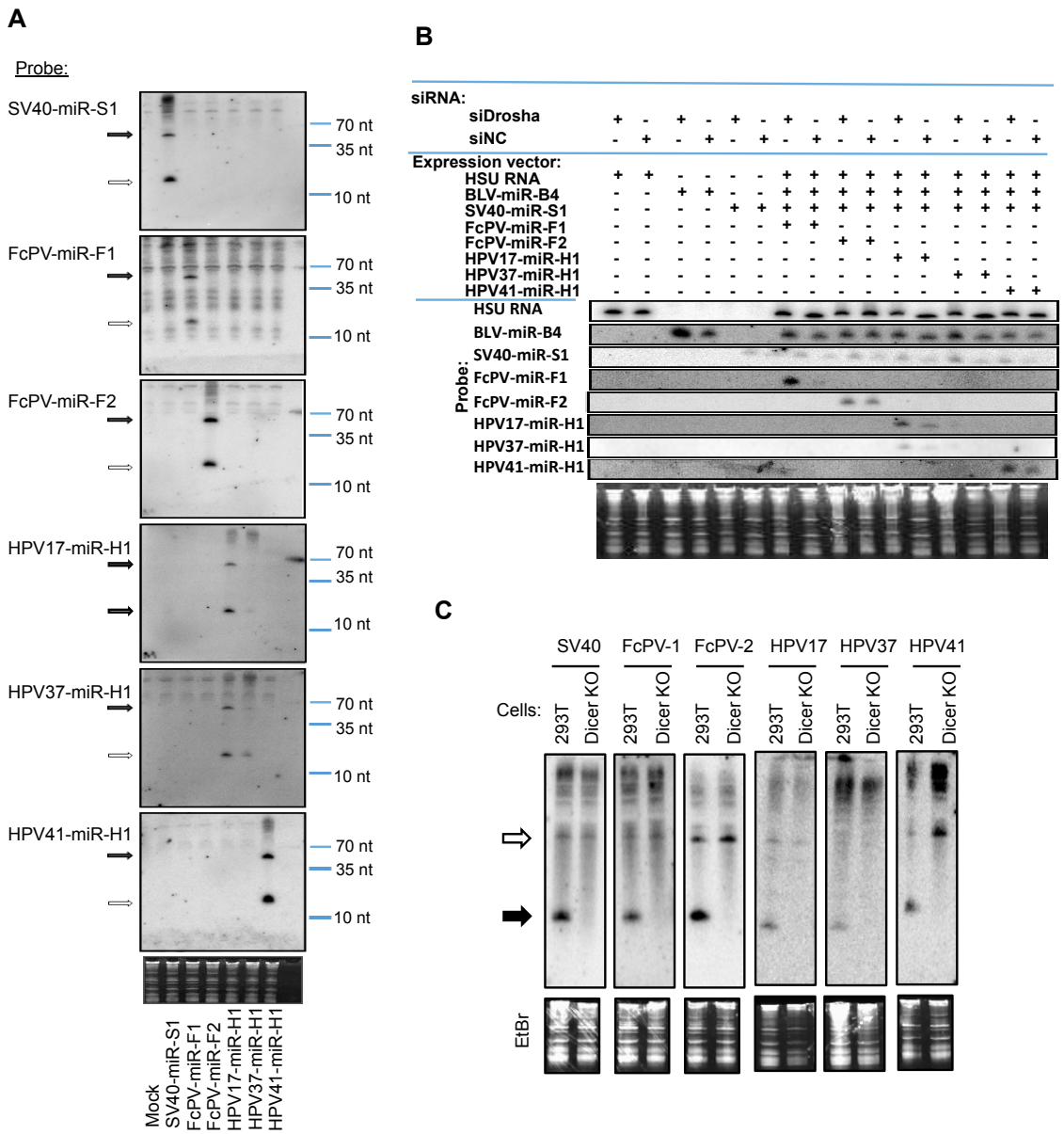


Figure 2.3: MiRNA candidates identified from MIDGE analysis are detected by Northern blot. Northern blot analysis of total RNA from HEK293T cells transfected with indicated miRNA or putative miRNA expression vectors, with ethidium bromide stained low molecular weight RNA shown as a load control. Blot was probed first for control SV40 miRNA, then stripped and re-probed for each of the indicated miRNAs. Solid and outline arrows correspond to pre-miRNAs and mature miRNAs, respectively. B) Northern blot analysis of total RNA from HEK293T cells transfected with anti-Drosha siRNA or negative control (NC) siRNA, then re-transfected after 48 hours with respective siRNAs and indicated putative miRNA expression vectors, and total RNA was harvested after 48 hours. Ethidium bromide stained low molecular weight RNA shown as a load control. Blot was probed first for control SV40 miRNA, then stripped and re-probed for each of the indicated miRNAs. Membrane was additionally probed for HSUR RNA as a transfection control, and Drosha-independent BLV miRNA. Normalized to HSUR RNA, all indicated miRNAs are detected at higher intensity in negative control treated cells than Drosha knockout cells, suggesting canonical dependence on Drosha processing. Note: due to sequence similarity between putative HPV-derived miRNAs, cross-reactivity is seen in HPV17-miR-H1 lanes and HPV37-miR-H1 lanes. C) Northern blot analysis of RNA from either HEK293T or Dicer KO (NoDice Ref) cells transfected with the indicated miRNA expression vectors. The numbers below each blot correspond to the ratio of mature miRNA signal to pre-miRNA signal. This figure represents a single membrane that was probed, then stripped and reprobed with the indicated oligonucleotide probes (B5, SV40, FcPV-1, FcPV-2, HPV41 probes) and a second membrane for HPV17 and HPV37. Ethidium bromide stained low molecular weight RNA is shown as a load control

Sample	FcPV1 Avg Ct +/- SD, N =	Let7a Avg Ct +/- SD, N =	Viral miRNA Present?	FcPV2 Avg Ct +/- SD N =	Let7a Avg Ct +/- SD, N =	Viral miRNA Present?
CF229-Chest	ND N = 2	33.5 +/- 2.1, N=2	NO	ND N = 3	34.9 +/- 0.7 N = 3	NO
CF229-Foot	30.3 +/- 1 N = 2	32 N = 1	YES	30.9 +/- 0.5 N = 3	33.9 +/- 0.3 N = 3	YES
CF151-Chest	ND N = 3	30.2 +/- 2 N = 3	NO	ND N = 3	29.3 +/- 2.2 N = 3	NO
CF151-Foot	33.8 +/- 0.4 N = 3	26.8 +/- 0.7 N = 3	YES	ND N = 3	30.8 +/- 2.8 N = 3	YES
CF180-Chest	ND, N = 1	34.0, N = 1	NO			
CF180-Foot	31.0, N = 1	28.8, N = 1	YES			

Table 2.1: FcPV miRNAs are found *in vivo* during viral infection: RNA was isolated from chaffinch tissues (Chest or Foot) and measured for the presence of each FcPV miRNA, as well as for the endogenous miRNA Let7a via Taqman qRT-PCR. Ct values are shown with Standard Deviation (SD), as well as the number of repetitions (N).

PV miRNAs and select candidates can directly regulate transcripts corresponding to the early viral genomic region.

Previous studies in the small DNA virus polyomavirus (PyV) family demonstrated that PyV miRNAs directly regulate early viral transcripts (8, 24). Furthermore, Bandicoot Papillomatosis Carcionomatosis Viruses 1 & 2 (BPCVs 1&2), hybrid viruses that have PyV-like early genes and genomic organization but PV-like capsid genes, also regulate early viral transcripts via viral miRNAs. Therefore, we performed bioinformatic analysis to examine the possibility that PV miRNAs could regulate early viral gene expression. To identify putative viral target transcripts, we identified seed complementary sites of 7 or more nucleotides for each PV miRNA in its respective genome. This analysis revealed 3 putative target sites for the five PV miRNAs/highly probable candidates (Figure 2D). On average this is in line with what would be predicted by chance for random 7-mers. Notably, both FcPV and HPV41 had candidate viral miRNA docking sites in the E1/E2 regions of their genomes, which has previously been demonstrated to include transcripts regulated by a host miRNA for HPV31 (12). We therefore tested these possible PV mRNA docking sites by engineering chimeric luciferase reporters containing the entire E1/E2 region of each respective genome. Co-transfection of these reporter plasmids with individual PV miRNA expression vectors revealed that E1 reporters for the FCPV and HPV41 genomes display significantly less expression in the presence of their respective

viral miRNAs (Figure 4B-E). Co-transfection of FcPV miR-F1 alone reduced expression of the FcPV E1/E2 reporter and this regulation was enhanced by co-transfecting plasmids expressing both FcPV miRs-F1 and F2 (Figures 4 B-D). Co-transfection of the HPV41 E1/E2 reporter and miRNA expression vectors demonstrated a significant reduction in luciferase expression. Because there were multiple predicted viral miRNA docking sites in both the FcPV and HPV41 E1/E2 regions, we engineered chimeric reporters that contained portions of the E1/E2 genomic region encompassing only a single predicted site. Co-transfection of the respective viral miRNA with these reporters resulted in a significant reduction in expression for both the E1 and E2 docking sites in both FcPV and the E1 site in HPV41. Importantly, negative control reporters containing 3 nucleotide mutations in the seed complement region of each predicted PV miRNA docking site alleviated the regulation that we observed (Figure 4B-E). These results demonstrate that FcPV and HPV41 miRNAs/candidates are able to directly regulate transcripts corresponding to the PV early genomic region.

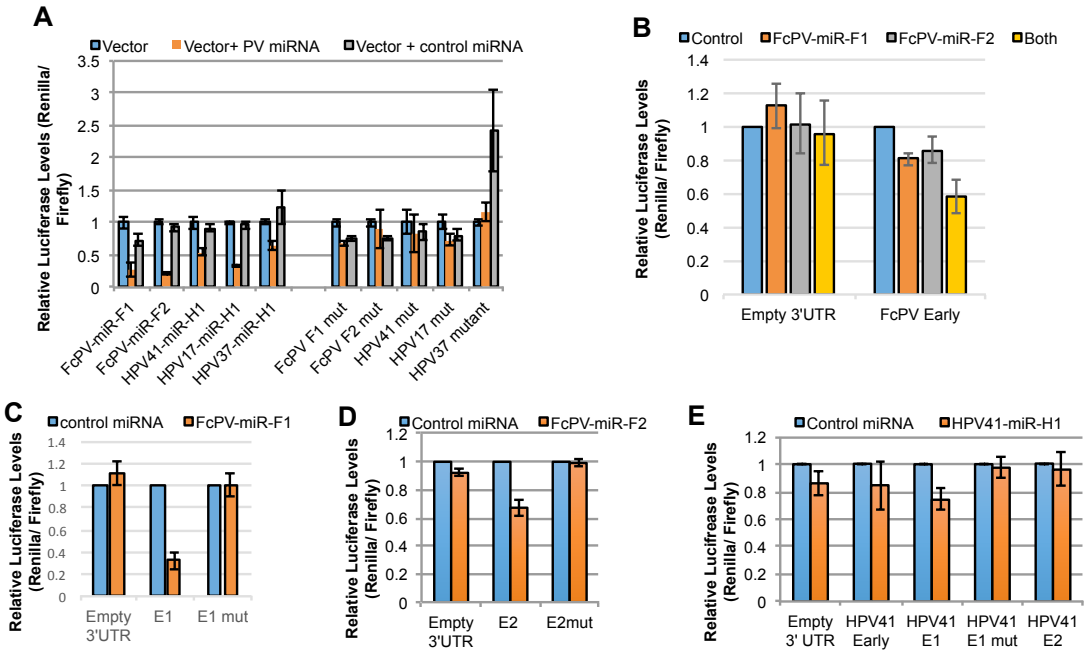


Figure 2.4: PV-encoded miRNAs are active in RISC and can regulate sequences in early genes: A-E) RISC reporter assays for the PV-encoded miRNAs. **A)** HEK293 cells were co-transfected with a firefly luciferase reporter and Renilla luciferase reporter with either perfectly complementary sequence matches for each indicated miRNA or its respective seed mutant, in addition to either empty vector, PV miRNA-expression vector, or control miRNA expression vector. Average Renilla luciferase activity relative to firefly luciferase normalized to empty vector controls is shown, N=3. **B)** HEK293T cells were co-transfected with either control or indicated PV miRNA expression vector and both control firefly luciferase reporter and Renilla luciferase based reporter plasmids with vector UTR (Empty) or FcPV genomic DNA containing both putatitve miRNA sites (FcPV Early) N=4. **C)** HEK293T cells were co-transfected with either the SV40 miRNA expression vector (Control) or the indicated PV miRNA expression vector and both control firefly luciferase reporter and Renilla luciferase based reporter plasmids with vector UTR (Empty) or FcPV E1 genomic sequence (E1 rLuc) or the seed sequence mutant (E1 mut), N=3 **D)** FcPV E2 genomic sequence (E2 rLuc) or the seed sequence mutant (E2 mut), N=12 **E)** HPV41 genomic DNA containing both putatitve miRNA sites (HPV41 Early), the site in E1 (HPV41 E1), the seed sequence mutant (E1 mut) or the site in E2 (HPV41 E2). N=7 The average Renilla luciferase activity normalized to firefly luciferase activity is shown. Error bars indicate Standard Error.

High risk PV types do not encode canonical miRNAs.

Our miDGE results provided more than 99% coverage in our RNA and DNA libraries for 65 PV genomes. Of these, only HPVs 17, 37, 41 & FcPV and gave rise to miRNAs/highly probable candidates. The only PVs that had small RNA:DNA read ratios consistent with miRNA size in the same range as our positive controls were HPV41,? HPV17, HPV37, and FcPV. These results strongly suggest that many PVs, including the high risk types 16 & 18 that were previously suggested to give rise to miRNAs do not encode canonical miRNAs. Therefore, to further evaluate whether high risk PVs code for miRNAs, we analyzed the transcriptomes of 200 cervical carcinomas deposited in The Cancer Genome Atlas (TCGA). We limited our analysis to those lesions that had sufficient (at least 100% RNA) coverage of a particular PV genome, which included HPVs 16, 18, 31 & 5. We then identified small RNA reads from these viruses and determined: 1) if they are enriched in any discreet regions of the genome, and 2) if they mapped to possible predicted hairpin secondary structures that are hallmarks of true pre-miRNAs. Although, we detected small RNAs in the appropriate size class of miRNAs derived from these transcriptomes, miRNA-size-class RNAs were not enriched over other sizes, nor did they map to probable predicted pre-miRNA structures (data not shown). Thus, despite having transcripts that span the entire viral genome, tumors associated with high-risk PV types do not express canonical miRNAs. Because it

was previously reported that transformed lines gave rise to HPV16 viral miRNAs (48, 53) we analyzed the deposited transcriptomic datasets from these studies for genomic regions with enrichment of miRNA-sized RNAs over other sizes of RNA. Similar to the PV-associated tumors, the HPV small RNAs present in these cell-line-derived libraries are not enriched for miRNA size classes. These results suggest that the small RNAs in these studies are degradation products derived from turnover of longer PV RNAs. Combining these findings with our miDGE results, we conclude that the high risk PVs (16, 18, & 31) do not express miRNAs.

DISCUSSION

Members of diverse virus families express miRNAs. These include the herpesviruses, polyomaviruses, anelloviruses and retroviruses (9, 36, 37, 38, 66). Notably, all of these viruses undergo persistent infection, have access to the nucleus where key miRNA processing machinery resides and are exclusively DNA viruses or have a DNA component to their lifecycle. Viruses with a persistent component to their life cycle may especially benefit from the typically subtle regulation afforded by miRNAs. Based on these characteristics, at least some members of the PV family would be expected to encode miRNAs. Yet, until now, no widely accepted PV miRNAs are known. Here we report the first high confidence papillomavirus-encoded miRNAs from a minor subset of PVs.

We identify PV miRNAs from the bird *Fringilla colebs* and high-confidence candidate miRNAs the human PVs 17, 37, and 41. These all display the hallmarks of canonical miRNAs, including being processed by Dicer and Drosha and being highly active in RISC. These RNAs derive from three divergent clades of PV. HPVs 17 and 37 are in the beta clade, some members of which have been proposed to have a role in skin cancers . The pre-miRNA hairpin region of the E2 locus for these viruses appears to have evolved in a common ancestor of these viruses, and may be shared with closely related HPVs15 & 80, which were not included in our library (data not shown). HPV41 is notable in that it is the sole member of the Nu clade and is one of the few PVs that have starkly different locations in a phylogenetic tree, depending if the trees are built upon late (L1) or the early proteins (E1). This implies HPV41 is likely a hybrid recombinant virus and may help to explain its atypical ability to encode a miRNA. FcPV is only distantly related to human PVs and is found in obvious highly keratinized large hyperplastic lesions afflicting the European Chaffinch songbird, *Fringilla colebs*. This allowed us to obtain RNA from PV-associated diseased tissue, which confirmed that FcPV miRNAs are expressed *in vivo* (Table 1) and confirmed their status as bona fide viral miRNAs.

Both the FcPV and HPV41 pre-miRNAs are found in non-protein-coding genomic locations, a common feature of miRNAs. One of the FcPV miRNA loci and the HPV41 miRNA locus are both found in a similar location downstream of the late genes past the likely late poly A cleavage site. In contrast, the HPV17

and 37 miRNAs are found in an identical genomic region overlapping and in the same transcriptional orientation as the L2 locus. Although it is atypical for a pre-miRNA gene to overlap a protein-coding gene, a similar genomic arrangement is observed for miR-K12 and the KaposinA gene in KSHV (25, 46, 47). Drosha can suppress Kaposin expression in latent infection, but its steady state levels and consequent ability to regulate Kaposin levels decrease during stress and at late times of lytic infection (47). Therefore, it is conceivable that HPVs 17 & 37 utilize Drosha to aid in controlling the expression of L2. HPV17 and 37 are the only PV miRNAs/candidates that share a high degree of sequence identity (91% 3P, 91% 5P), consistent with their being derived from virus types that are closely related. Except for HPVs17 and 37, there is no obvious relationship between the miRNA-positive PVs that might account for why they would preferentially encode miRNAs.

miDGE covered at least 500,000 nucleotide of sequence space (the sum of viral genomes and bacterial plasmid sequences covered in our libraries) and did not call any miRNAs other than the the positive controls and the five PV miRNAs that we validated. Therefore, we conclude that miDGE has an intrinsic low false positive rate. These findings are consistent with the current understanding that pre-miRNA hairpins have specific structural features that are required for efficient processing (9). In contrast, only a subset of PV genomes (n= 48) had 100% coverage in both our DNA and RNA libraries. Therefore, we can only make negative conclusions on this limited set of PVs. Our results show that

44/48PV genomes with complete DNA and RNA coverage lack the ability to efficiently give rise to canonical miRNAs. This list includes the high-risk PVs 16. 18 & 31, which our further analysis demonstrated do not encode miRNAs in tumor or cancer cell line settings . Although we acknowledge that our findings are skewed toward the alpha clade human PVs, we conclude that numerous PVs lack the ability to encode their own miRNAs.

What are the functions of the PV miRNAs? Our bioinformatic analysis and reporter assays (Figure 4) suggest that one function of PV miRNAs is to regulate viral gene expression. This notion is consistent with the known function of other viral mRNAs, especially those derived from the PyVs and BPCVs. Since our results suggest that many PVs do not encode miRNAs, this raises the question of if/how such viruses fine-tune their own gene expression. At least for one high risk PV, HPV31, the answer seems to be by co-opting host miRNAs. Gunasekharan et al. demonstrated that miR-145 is expressed at higher levels in undifferentiated versus differentiated keratinocytes, displaying an inverse pattern to HPV31 replication levels (30). miR-145 negatively regulates HPV31 genome replication and gene expression. Interestingly, miR-145 directly docks to and regulates the E1/E2 transcripts in a genomic region similar to that we have uncovered for FcPV and HPV41 miRNAs.

Our results demonstrate that miDGE can be a fruitful approach when applied to numerous viruses. However, while miDGE allowed us to make conclusions for ~60 PV genomes, some genomes that we intended to include in

our miDGE procedure were under-represented or not represented in the final libraries. Although we believe one major reason for this discrepancy is due to incorrect PVgenomic plasmids included in our original library pools, library coverage should be optimized in future applications of miDGE. Since miRNAs are generally stable, using miDGE could be used to identify biomarkers for gene expression of un-culturable pathogens. In addition to pathogen genomes that cannot be grown in culture, miDGE may have utility for identifying miRNAs expressed in only a few rare cells of an organism. For example, miR-Lys6 is only expressed in fewer than 10 cells in C. elegans and has been missed by most standard miRNA biochemical identification procedures .It is likely that similar miRNAs exist in complex multicellular organisms and these could be identified using miDGE.

In summary, we have developed wet bench technology that can identify miRNAs from genomes in which complete transcriptomes are not readily available. Our approach uncovered five new PV miRNAs/highly probable candidates. As viral miRNAs often alter host targets conducive to infection, it will be interesting to determine host targets of these miRNAs. Furthermore, as our work lends further support to the role of miRNAs in control of the PV life cycle, it will be important to determine if variability in miRNA expression or activity can contribute to the differences in tropism and pathogenesis associated with the various PV types.

Chapter 4: Materials and Methods

Materials and Methods: SERPINA1 3'UTR variation does not contribute to Alpha-1-Antitrypsin Deficiency disease severity

Cell culture

HEK293, HepG2, and Huh--7 cells were obtained from ATCC and maintained in DMEM supplemented with 10% (vol/vol) FBS and Pen/Strep (Cellgro). All cells were grown at 37C in the presence of 5% CO2.

Plasmids

The wild-type AAT UTR sequence was amplified via PCR out of genomic DNA extracted from cultured cells using 3'UTR primer set (Appendix Table 1.1). The amplicons were then cloned into the Xho1/Not1 sites of pSICHECK 2 (Promega). The polyA mutant vector was generated by PCR-directed mutagenesis using polyA mutant sense and polyA mutant antisense primers (Appendix Table 1.1). The polyA deletion mutant vector was generated by PCR-directed mutagenesis using 3'UTR deletion mutant sense and 3'UTR deletion mutant antisense primers (Appendix Table 1.1) Nano-Luc plasmid was obtained from Promega.

Knockdown of SERPINA1

HepG2 or Huh-7 cells (12-well format, 70% confluency) were reverse transfected with Nano-Luc plasmid (Promega). Forty-eight hours later, cells were pooled and reverse transfected (12-well format, 70% confluency) with 20-pmol of either negative control siRNA (scrambled; Sigma-Aldrich) or SERPINA1-specific siRNAs (Sigma-Aldrich; part #: AM16708; siRNA ID#: 105842). Forty-eight hours later, cell culture supernatant was harvested (1 mL/ well) and total RNA was isolated using PIG-B (74). RNA was treated with DNase I to remove genomic DNA and re-isolated using NaOAC extraction. Poly-adenylated transcripts were then reverse transcribed by Superscript III enzyme using poly-d(T) primer, and SerpinA1 cDNA was quantified relative to GAPDH control by quantitative Real-Time PCR using SYBR Green Master Mix and StepOne Real-Time PCR system from Applied Biosystems. Transfections were performed in triplicate. Secreted Nano-Luc was measured by Nano-Glo assay forty-eight hours post-transfection using GloMax® 96 Microplate Luminometer (Promega). Knockdown was confirmed by ELISA measuring secreted SERPINA1 gene product from cell culture supernatant using the SynergyMx Microplate Reader from BioTek. Data was normalized to negative control siRNAs and normalized for overall secretion levels by NanoLuc levels.

miRNA Transfection

HepG2 or Huh-7 cells (12-well format, 70% confluency) were reverse transfected with Nano-Luc plasmid (Promega). Forty-eight hours later, cells were

pooled and reverse transfected (12-well format, 70% confluency) with 20-pmol of either negative control siRNA (scrambled; Sigma-Aldrich) or SERPINA1-specific siRNAs (Sigma-Aldrich; part #: AM16708; siRNA ID#: 105842) as controls for SERPINA1 knockdown, or with either negative control miRNA mimic (scrambled, Sigma-Aldrich) or one of fifty liver-specific miRNA mimics (Sigma-Aldrich, Supplementary Table). Forty-eight hours later, cell culture supernatant was harvested (1 mL/ well) and total RNA was isolated using PIG-B (74). SiRNA-mediated knockdown was assayed as above to confirm transfection efficiency. Secreted Nano-Luc was measured by Nano-Glo assay forty-eight hours post-transfection using GloMax® 96 Microplate Luminometer (Promega). Relative AAT levels were assayed by ELISA measuring AAT levels from cell culture supernatant using the SynergyMx Microplate Reader from BioTek Data was normalized to negative control siRNAs and normalized for overall secretion levels by NanoLuc levels. MiRNA mimics that scored 20% above or below negative control levels in both cells lines were re-screened in triplicate in both cell lines.

Reducing miRNA Expression Assay

To determine the effect of miRNA under-expression on secreted AAT, Huh-7 cells were infected (MOI of 500) with either control adenoviral vector AdGFP or a VP55 vector AdGFPVP55, or transfected with either negative control

siRNA (Sigma-Aldrich) or SERPINA1 siRNA (Sigma-Aldrich). Forty-eight hours post-infection or post-transfection, cell culture supernatant was collected and total RNA was isolated using PIG-B (74). Northern blots were performed as described in McClure et al., 2011. Briefly, RNA was isolated using PIG-B (74), fractionated on 15% PAGE-UREA gel, and transferred to Amersham Hybond – N+ membrane (GE Healthcare). The membrane was then probed in ExpressHyb hybridization solution (Clontech, CA) with indicated DNA oligos (Integrated DNA technologies; Supplemental Table 2) radio-labeled with [γ-32P]-ATP (6000 Ci/mmol) by T4 Polynucleotide Kinase (New England Biolabs). Membranes were exposed to a storage phosphor screen (GE Healthcare) and visualized with a Typhoon biomolecular imager (GE Healthcare). Upon confirmation of miRNA depletion, cell culture supernatants were assayed for secreted AAT levels using Abcam ELISA kit.

3'UTR-Mediated Regulation Assay:

For analysis of 3'UTR-mediated regulation in liver and non-liver cells, HEK293 or HepG2 cells (12-well format, 70% confluent) were transfected with 50 ng of the indicated Renilla Luciferase based reporter constructs (pSICHECK2, Promega) using Lipofectamine 2000 according to the manufacturer's instructions. Transfections were carried out in triplicate for each transfection, as a technical replicate. Forty-eight hours post transfection, lysates were harvested

and dual luciferase assay was performed. Raw renilla/firefly ratios were normalized to pSICHECK2 parental vector for each data set.

Alternative Poly-Adenylation Assay:

For analysis of alternative UTR mediated regulation of SERPINA1, HEK293 cells (12-well format, 70% confluent) were transfected with 50 ng of the indicated Renilla Luciferase based reporter constructs (pSICHECK2, Promega) using Lipofectamine 2000 according to the manufacturer's instructions. Transfections were carried out in triplicate for each transfection, as a technical replicate. Forty-eight hours post transfection, lysates were harvested and dual luciferase assay was performed. Raw renilla/firefly ratios were normalized to pSICHECK2 parental vector for each data set. Production of alternative length 3'UTRs was verified by RT-PCR using primers to detect the full-length (1.7 Kb) UTR in varying amounts, visualized by ethidium bromide staining of agarose gels.

Patient 3'UTR Variation Detection:

To determine whether or not patients with varying degrees of A1AD disease could be grouped by single nucleotide variations in the SERPINA1 3'UTR, the SerpinA1 3'UTR was amplified out of genomic DNA of each of 250 patients by PCR. Amplification of a single band per patient was verified by ethidium bromide staining of agarose gels, and each patient UTR was gel

extracted using Zymoclean DNA Gel Extraction kit. Each UTR was sequenced by Sanger sequencing, and chromatograms were analyzed using Geneius to detect variations in sequencing results. Additionally, patients were selected at random for resequencing to confirm initial results and sex of these patients were confirmed by determining the presence or absence of a Y chromosome-specific gene. Finally regions containing known common SNP rs4149057 were amplified from genomic DNA and sequenced for these patients to confirm that SNPs could indeed be detected in expected ratios in non-disease related gene regions.

Materials and Methods: Identification of virus-encoded miRNAs in divergent Papillomaviruses

miDGE:

For JMRV, a 20kb genomic region was PCR-amplified using ??? as template. We then sonicated to enrich for DNA fragments displaying a peak in size at approximately 1000 nucleotides. For the PVs, plasmids containing cloned PV genomes were collected from various labs. These plasmids were then grouped into sub-collections and for each collection of plasmids, separate libraries were made from digesting them with different 4-base pair cutter restriction enzymes (In addition, we also generated a library with mostly overlapping PV genomes that generated via sonication (as described above for JMRV). We then generated libraries in the pcDNA3.1 background, transfected HEK293T cells via lipofection, and harvested total RNA using PigB. RNA was

then size fractionated via excision from a 15% denaturing polyacrylamide gel enriching for RNA between approximately 10-35 nucletodies.

Plasmids and Cells:

HEK293T cells were obtained from ATCC and maintained in DMEM supplemented with 10% (vol/vol) FBS and Pen/Strep (Cellgro). HEK293T Dicer KO cells were obtained from the Cullen lab(6). All cells were grown at 37C in the presence of 5% CO2.

Plasmids containing the genomes of the papillomaviruses used in this study were obtained from labs indicated in Appendix Table 1.

The miRNA expression vectors were cloned through PCR amplification of the relevant portions of the viral genome (from the genomic plasmids), followed by restriction enzyme digestion and ligation. Specifically, the indicated regions of each viral genome listed in Appendix Table 1 were inserted into the XhoI/XbaI sites of pcDNA3.1 (Invitrogen). The SV40 miRNA expression construct is as previously described (2).Seed site mutant constructs altered the indicated positions 2, 3, and 5 of the seed site (FcPV mutant constructs), or positions 2, 3 (HPV41 mutant construct) .To make RISC reporters for each miRNA, sequences with two perfectly complementary binding sites for each miRNA with a 12 nucleotide spacer region were cloned into the Xho1/Xba1 sites of pcDNA3.1dsRLuc plasmid as listed. Respective mutants were made by mutating

three nucleotides in the seed sequence of each binding site, also listed in Appendix Table 2.2.

RISC Activity Assay:

HEK293T cells were split and plated onto twelve well dishes so that they were approximately 70% confluent the following day. These plates were then co-transfected with five ng of the indicated Renilla Luciferase based reporter constructs (pcDNA3.1dsRluc (3)), five ng of Firefly Luciferase reporter (pcDNA3.1dsLuc2CP (3)) and one ug of either a control miRNA expression construct (the SV40 miRNA expression construct) or the indicated miRNA expression vector using Lipofectamine 2000 according to the manufacturer's instructions. Transfections were carried out in triplicate for each transfection, as a technical replicate. Forty-eight hours later, cells were harvested with 100 uL of 1X Lysis buffer from the Dual-Glo Luciferase Assay System (Promega). 5 uL of lysate from each well was then analyzed in duplicate for Renilla and Firefly luciferase activity with a Luminoskan Ascent luminometer (Thermo Electronic).These experiments were then performed for at least 3 biological replicates (new transfections on separate days), with the exact number noted in the individual figure legend. Data was analyzed by dividing the Renilla luciferase activity value by the Firefly luciferase activity value to obtain a Renilla/Firefly luciferase activity ratio. These ratios were then averaged between the two measures for each well of the twelve well dish. Then the averaged Renilla/Firefly

ratio was averaged for each of the technical triplicate wells, with the resulting average used to normalize the final values such that the control miRNA Renilla/Firefly ratio was set to 1. The one-sided Student's t-test was used to assess the statistical significance of observed differences, with a P value of < 0.05 considered statistically significant.

miRNA Detection Assay:

HEK293T cells were transfected with the indicated miRNA expression constructs with Lipofectamine 2000 (Invitrogen). 48 hours post-transfection, total RNA was harvested with PIG-B and Northern blot analysis was performed as described in McClure et al (51). Briefly, RNA was transferred to Amersham Hybond –N+ membrane (GE Healthcare) and probed with DNA oligos indicated in Appendix Table 2.1.

Drosha Dependence Assay:

HEK293T cells were transfected with 20 nM Drosha siRNA or negative control siRNA using Lipofectamine RNAiMAX (Invitrogen). Forty-eight hours later these cells were pooled, replated to new wells and transfected with respective siRNAs and the indicated miRNA expression constructs with Lipofectamine 2000 (Invitrogen). Forty-eight hours later, total RNA was harvested with PIG-B and Northern blot analysis was performed as previously described 52. Northern blot membrane was probed with DNA oligos indicated in Appendix Table 2.1.

Dicer Dependence Assay:

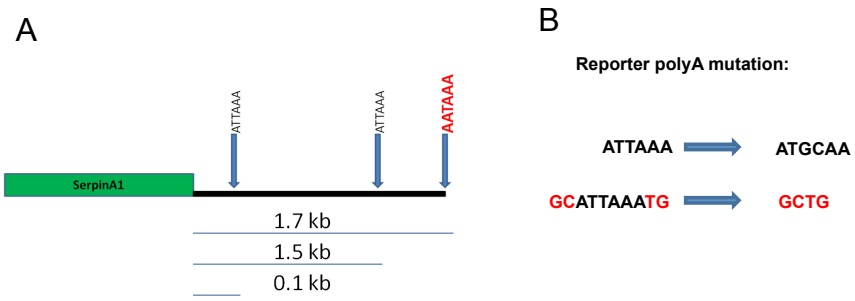
HEK293T and HEK293T Dicer KO cells (6) were transfected with the indicated miRNA expression vectors using Lipofectamine 2000 (Invitrogen). Forty-eight hours post-transfection, total RNA was harvested with PIG-B and Northern blot analysis was performed. Northern blot membrane was probed with DNA oligos indicated in Table X.1.

***In Vivo* miRNA Assay:**

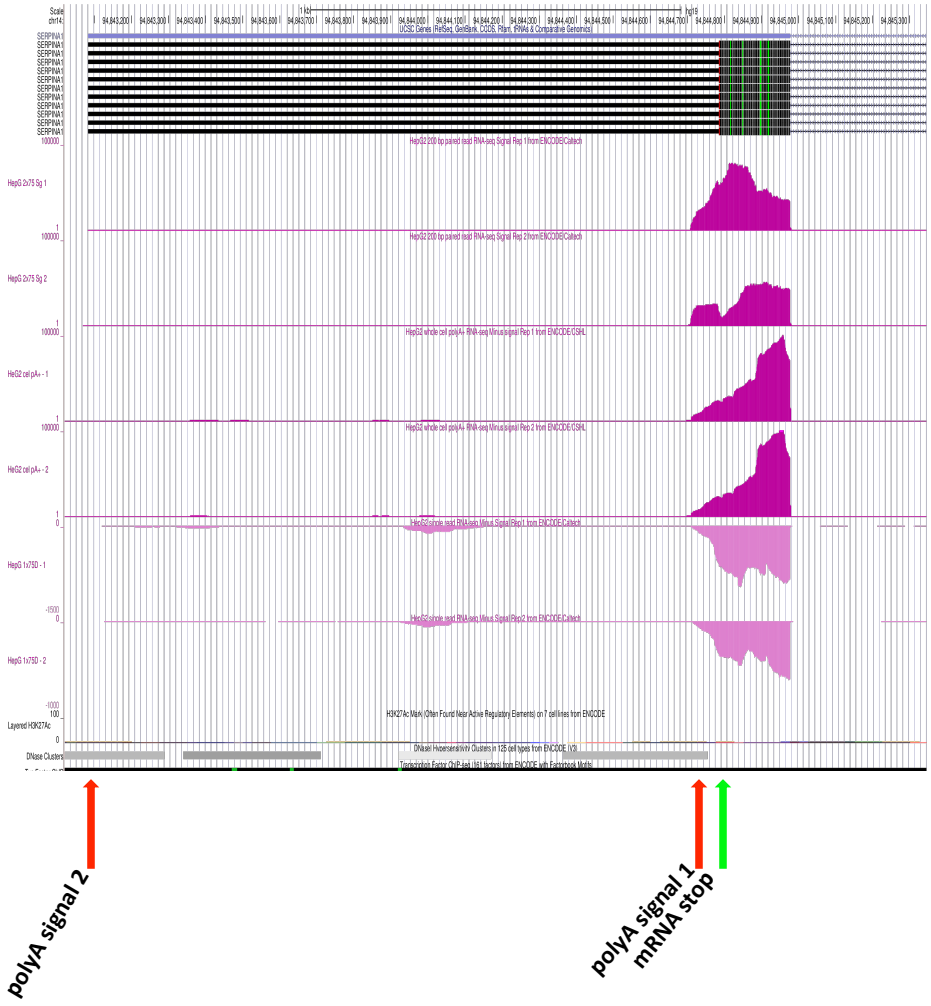
Total RNA was extracted from both foot and pectoral muscle tissue from chaffinches that displayed clinical symptoms of FcPV infection. This RNA was then analyzed using TaqMan miRNA quantitation kits that were designed to each FcPV miRNA or the host Let7a miRNA using the manufacturer's instructions, with the following exceptions. The RT-PCR reaction was performed with the Let7a and one of the two viral miRNA-specific primers. The product of this reaction was then split for individual qPCR reactions using probe/primer sets that were specific for either the viral or host miRNAs using StepOne Real-Time PCR system from Applied Biosystems.

Appendices

Appendix 1: SERPINA1 3'UTR variation does not contribute to Alpha-1-Antitrypsion Deficiency disease severity



Appendix Figure 1.1: SERPINA1 3'UTR Alternative poly-adenylation (A) SerpinA1 3' noncoding region contains multiple potential poly-A signal sequences. (B) Poly-A- signal sequence mutants constructed.



Appendix Figure 1.2: Data from ENCODE (67) showing abundance of reads several fold higher at proximal poly-A-signal sequence.

Appendix Table 1.1: Primer sets used for SERPINA1 3'UTR analysis

Luciferase reporter primers	
AAT 3'UTR F	5'-tagcCTCGAGtcccacccaaaaataactgc-3'
AAT 3'UTR R	5'-tagctctagaAGCCAGGGGAGGTACTTCAT-3'
AAT 3'UTR poly A mutant F	5'-gcaagaagggttgagctggt-3'
AAT 3'UTR poly A mutant R	5'-atgtcatccagggaggggg-3'
AAT 3'UTR deletion mutant F	5'-gaagggttgagctggtcc-3'
AAT 3'UTR deletion mutant R	5'-tgtcatccagggaggg-3'
InvertedF	5'-atcgctcgagAGCCAGGGGAGGTACTTCAT-3'
InvertedR	5'-atcggcggccgcTCCCACCCAAAAATAACTGC-3'
SNP sequencing primers	
rs4149057 F	5'-cttcatcttccgccatga-3'
rs4149057 R	5'-gatcccagggtaaagccaat-3'
T-Specific Y-linked primer F	5'-CGTCAGGTGAGCCGTTAGTT-3'
T-specific Y-linked primer R	5'-ACACATCCCACACCCATTCA-3'
qPCR primers	
SERPINA1 F	5'-gggcatcactaaggtcttca-3'
SERPINA1 R	5'-tgaagaggggagacttggta3'
GAPDH F	5'-acatcgctcagacaccatg-3'
GAPDH R	5'-tgtagttgaggtcaatgaaggg-3'
3'RACE primers	
RACE adapter	5'-GACTCGAGTCGACATCGTTTTTTTTTTTTTTTTTTTT3VN-3'
FWD primer	CAAGAGCTTCGTGGAGC
REV primer	GACTCGAGTCGACATCG

Appendix 2: Identification of virus-encoded miRNAs in divergent Papillomaviruses

Appendix Table 2.1: All papillomavirus genome plasmid included in miDGE analysis

Genome	Genus	Obtained From	Affiliation	Location
Human Papillomavirus 1	Mu	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 2	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 3	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 4	Gamma	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 5	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 6b	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 7	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 8	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 9	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 10	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 11	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 12	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 13	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 14D	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 15	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 16	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 17	Beta	M. Favre	Institut Pasteur	Paris, France

Human Papillomavirus 18	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 19	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 20	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 21	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 22	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 23	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 24	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 26	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 27	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 28	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 29	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 30	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 31	Alpha	Lorna Sammoury	Qiagen	Washington, DC
Human Papillomavirus 32	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 33	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 34	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 36	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 37	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 38	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 40	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 41	Nu	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 42	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 45	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany

Human Papillomavirus 47	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 48	Gamma	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 49	Beta	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 50	Gamma	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 51	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 52	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 53	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 54	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 56	Alpha	Lorna Sammoury	Qiagen	Washington, DC
Human Papillomavirus 57	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 58	Alpha	T. Matsukura	Kyoto University	Kyoto, Japan
Human Papillomavirus 59	Alpha	T. Matsukura	Kyoto University	Kyoto, Japan
Human Papillomavirus 61	Alpha	T. Matsukura	Kyoto University	Kyoto, Japan
Human Papillomavirus 62	Alpha	R. Burk	Albert Einstein College of Medicine	Bronx, New York
Human Papillomavirus 63	Mu	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 65	Gamma	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 66	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 67	Alpha	T. Matsukura	Kyoto University	Kyoto, Japan
Human Papillomavirus 68a	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 69	Alpha	T. Matsukura	Kyoto University	Kyoto, Japan
Human Papillomavirus 70	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 71	Alpha	T. Matsukura	Kyoto University	Kyoto, Japan
Human Papillomavirus 72	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany

Human Papillomavirus 74	Alpha	M. Favre	Institut Pasteur	Paris, France
Human Papillomavirus 75	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 76	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 81	Alpha	T. Matsukura	Kyoto University	Kyoto, Japan
Human Papillomavirus 82	Alpga	T. Matsukura	Kyoto University	Kyoto, Japan
Human Papillomavirus 92	Beta	O. Forslund	Lund University	Lund, Sweden
Human Papillomavirus 93	Beta	O. Forslund	Lund University	Lund, Sweden
Human Papillomavirus 94	Alpha	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 96	Beta	O. Forslund	Lund University	Lund, Sweden
Human Papillomavirus 98	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 99	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 100	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 104	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 105	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 107	Beta	O. Forslund	Lund University	Lund, Sweden
Human Papillomavirus 108	Gamma	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 109	Gamma	O. Forslund	Lund University	Lund, Sweden
Human Papillomavirus 110	Beta	O. Forslund	Lund University	Lund, Sweden
Human Papillomavirus 111	Beta	O. Forslund	Lund University	Lund, Sweden
Human Papillomavirus 112	Gamma	O. Forslund	Lund University	Lund, Sweden
Human Papillomavirus 113	Beta	E.-M. de Villiers	German Cancer Research Center	Heidelberg, Germany
Human Papillomavirus 114	Alpha	O. Forslund	Lund University	Lund, Sweden
Human Papillomavirus 117	Alpha	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany

Human Papillomavirus 118	Beta	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Human Papillomavirus 125	Alpha	A. Kovanda	Jožef Stefan Institute	Ljubljana, Slovenia
Human Papillomavirus 127	Gamma	C. Buck	National Institutes of Health	Bethesda, Maryland
Human Papillomavirus 128	Gamma	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Human Papillomavirus 129	Gamma	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Human Papillomavirus 130	Gamma	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Human Papillomavirus 131	Gamma	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Human Papillomavirus 132	Gamma	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Human Papillomavirus 133	Gamma	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Human Papillomavirus 134	Gamma	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Human Papillomavirus 148	Gamma	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Human Papillomavirus 149	Gamma	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Human Papillomavirus 150	Beta	A. Kovanda	Jožef Stefan Institute	Ljubljana, Slovenia
Human Papillomavirus 151	Beta	A. Kovanda	Jožef Stefan Institute	Ljubljana, Slovenia
Bandicoot PV 1		M. Bennett	Small Animal Specialist Hospital	North Ryde, Australia
Bovine Papillomavirus 1	Delta	C. Buck	National Institutes of Health	Bethesda, Maryland
Camelus dromedarius Papillomavirus 1	Delta	O. Forslund	Lund University	Lund, Sweden
Camelus dromedarius Papillomavirus 2	Delta	O. Forslund	Lund University	Lund, Sweden
Delphinius delphis Papillomavirus	Upsilon	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany

European hedgehog Papillomavirus	Dyo	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Felis domesticus Papillomavirus 1	Lambda	Van Ranst	Rega Institute for Medical Research	Leuven, Belgium
Fringilla coelebs Papillomavirus 1	Eta	R. Burk	Albert Einstein College of Medicine	Bronx, New York
Lynx rufus Papillomavirus 1	Lambda	Van Ranst	Rega Institute for Medical Research	Leuven, Belgium
Mus musculus Papillomavirus 1	Pi	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Panthera leo persica Papillomavirus 1	Lambda	Van Ranst	Rega Institute for Medical Research	Leuven, Belgium
Phocoena phocoena Papillomavirus 1	Omikron	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Psittacus erithacus timneh Papillomavirus 1	Theta	R. Burk	Albert Einstein College of Medicine	Bronx, New York
Puma concolor Papillomavirus 1	Lambda	Van Ranst	Rega Institute for Medical Research	Leuven, Belgium
Rattus norvegicus Papillomavirus 1	Pi	I. Nindl	Charité – Universitätsmedizin Berlin	Berlin, Germany
Rhesus monkey Papillomavirus	Alpha	C. Buck	National Institutes of Health	Bethesda, Maryland
Rousettus gegypiacus Papillomavirus 1	Psi	A. Rector	Rega Institute for Medical Research	Leuven, Belgium
Sus scrofa Papillomavirus 1	Dyodelta	A. Rector	Rega Institute for Medical Research	Leuven, Belgium
Tursiops truncatus Papillomavirus 1	Upsilon	A. Rector	Rega Institute for Medical Research	Leuven, Belgium
Unicia uncia Papillomavirus 1	Lambda	Van Ranst	Rega Institute for Medical Research	Leuven, Belgium
Ursus maritimus Papillomavirus 1	Omega	A. Rector	Rega Institute for Medical Research	Leuven, Belgium

Appendix Table 2.2: Primers and oligos used in PV miRNA analysis

PV miRNA sequence	
FcPV miR-F1 perfect complement	5'-AAATTCCGGAGAGGGAACCGATTactgactgAAATTCCGGAGAGGGAACCGATT-3'
FcPV miR-F1 mutant site	5'-AAATTCCGGAGAGGGAAGGCATTatcgatcgatcgAAATTCCGGAGAGGGAAGGCATT-3'
FcPV miR-F2 perfect complement	5'-AGTATCTACTACCCGACCATATatcgatcgatcgAGTATCTACTACCCGACCATAT-3'
FcPV miR-F2 mutant site	5'-AGTATCTACTACCCCAGGATATatcgatcgatcgAGTATCTACTACCCCAGGATAT-3'
HPV17 miR-H1 perfect complement	5'-TCTTACAGGGGGACTACCATCAatcgatcgatcgTCTTACAGGGGGACTACCATGA-3'
HPV17 miR-H1 mutant site	5'-TCTTACAGGGGGACTAGGATCAatcgatcgatcgTCTTACAGGGGGACTAGGATCA-3'
HPV37 miR-H1 perfect complement	5'-TGATCCTAGTCCCCCTGTGCGACAGGGTACCGATTGATCCTAGTCCCCCTGTGCGA-3'
HPV37 miR-H1 mutant site	5'-TGATCCTAGTCCCGGTCTGCGACAGGGTACCGATTGATCCTAGTCCCGGTCTGCGA-3'
HPV41 miR-H1 perfect complement	5'-CACATGGACCGTCGAGGACACCatcgatcgatcgCACATGGACCGTCGAGGACACC-3'
HPV41 miR-H1 mutant site	5'-CACATGGACCGTCGAGCAGAGCatcgatcgatcgCACATGGACCGTCGAGCAGAGC-3'
Northern blot oligos	
FcPV miR-F1 probe	5'- AATCGGTTCCCTCTCCGGAATTT -3'
FcPV miR-F2 probe	5'- ATATGGTCGGGTAGTAGATACT -3'
HPV17 miR-H1 probe	5′ - TCTTACAGGGGGACTAGGATCA-3′
HPV37 miR-H1 probe	5′ - TCGCACAGGGGGACTAGGATCA-3′
HPV41 miR-H1 probe	5′ -GGTGTCCTCGACGGTCCATGTG-3′
SV40 miR-S1 probe	5′ -GGCATGAAACA GGCA-3′
BLV miR-B4 probe	5'-AAAGGCGCAGAGACTGTGGTGCTA-3'
HSURNA probe	5'-AGAGTAACTCTCTGGCTGTGGGC-3'

PV E gene reporters	
FcPV E region	nucleotides 2401-3870
FcPV E1 region	nucleotides 2003-2848
FcPV E2 region	nucleotides 3501-3890
HPV 41 E region	nucleotides 1567-3027
HPV 41 E1 region	nucleotides 1576-1982
HPV 41 E2 region	nucleotides 2683-3027
HPV17 E region	
HPV37 E1 region	nucleotides 899-2728
HPV37 E2 region	nucleotides 2670-4034

Bibliography

1. Axtell, M. J. "Vive La Différence: Biogenesis and Evolution of MicroRNAs in Plants and Animals." Genome Biology. U.S. National Library of Medicine, 28 Apr. 2011. Web. 27 July 2017.
2. Backes, S., J. S. Shapiro, L. R. Sabin, A. M. Pham, I. Reyes, B. Moss, S. Cherry, and B. R. TenOever. "Degradation of Host MicroRNAs by Poxvirus Poly(A) Polymerase Reveals Terminal RNA Methylation as a Protective Antiviral Mechanism." Cell Host & Microbe. U.S. National Library of Medicine, 16 Aug. 2012. Web. 27 July 2017.
3. Backes, Simone, Ryan A. Langlois, Sonja Schmid, Andrew Varble, Jaehee V. Shim, David Sachs, and Benjamin R. TenOever. "The Mammalian Response to Virus Infection Is Independent of Small RNA Silencing." Cell Reports. U.S. National Library of Medicine, 10 July 2014. Web. 27 July 2017.
4. Bartel, D. P. "MicroRNAs: Target Recognition and Regulatory Functions." Cell. U.S. National Library of Medicine, 23 Jan. 2009. Web. 27 July 2017.
5. Beaudoing, E., S. Freier, J. R. Wyatt, J. M. Claverie, and D. Gautheret. "Patterns of Variant Polyadenylation Signal Usage in Human Genes." Genome Research. U.S. National Library of Medicine, July 2000. Web. 27 July 2017.
6. Bogerd, H. P., A. W. Whisnant, E. M. Kennedy, O. Flores, and B. R. Cullen. "Derivation and Characterization of Dicer- and MicroRNA-deficient Human Cells." RNA (New York, N.Y.). U.S. National Library of Medicine, June 2014. Web. 27 July 2017.
7. Bogerd, H. P., H. W. Karnowski, X. Cai, J. Shin, M. Pohlers, and B. R. Cullen. "A Mammalian Herpesvirus Uses Noncanonical Expression and Processing Mechanisms to Generate Viral MicroRNAs." Molecular Cell. U.S. National Library of Medicine, 15 Jan. 2010. Web. 27 July 2017.
8. Broekema, N. M., and M. J. Imperiale. "MiRNA Regulation of BK Polyomavirus Replication during Early Infection." *Proceedings of the National Academy of Sciences of the United States of America.* U.S. National Library of Medicine, 14 May 2013. Web. 28 July 2017.
9. Burke, J. M., C. R. Bass, R. P. Kincaid, and C. S. Sullivan. "Identification of Tri-phosphatase Activity in the Biogenesis of Retroviral MicroRNAs and RNAP III-generated ShRNAs." Nucleic Acids Research.U.S. National Library of Medicine, 16 Dec. 2014. Web. 27 July 2017.

10. Burke, J. M., D. P. Kelenis, R. P. Kincaid, and C. S. Sullivan. "A Central Role for the Primary MicroRNA Stem in Guiding the Position and Efficiency of Drosha Processing of a Viral Pri-miRNA." RNA (New York, N.Y.). U.S. National Library of Medicine, July 2014. Web. 28 July 2017
11. Byth, B. C., G. D. Billingsley, and D. W. Cox. "Physical and Genetic Mapping of the Serpin Gene Cluster at 14q32.1: Allelic Association and a Unique Haplotype Associated with Alpha 1-antitrypsin Deficiency." American Journal of Human Genetics. U.S. National Library of Medicine, July 1994. Web. 27 July 2017.
12. Cai, Xuezhong, Gu Li, Laimonis A. Laimins, and Bryan R. Cullen. "Human Papillomavirus Genotype 31 Does Not Express Detectable MicroRNA Levels during Latent or Productive Virus Replication." Journal of Virology. American Society for Microbiology, Nov. 2006. Web. 27 July 2017.
13. Cech, T. R., and J. A. Steitz. "The Noncoding RNA Revolution-trashing Old Rules to Forge New Ones." Cell. U.S. National Library of Medicine, 27 Mar. 2014. Web. 27 July 2017.
14. Chen, Xian-Ming. "MicroRNA Signatures in Liver Diseases." World Journal of Gastroenterology : WJG. The WJG Press and Baishideng, 14 Apr. 2009. Web. 27 July 2017.
15. Cox, D. W., S. L. Woo, and T. Mansfield. "DNA Restriction Fragments Associated with Alpha 1-antitrypsin Indicate a Single Origin for Deficiency Allele PI Z." Nature. U.S. National Library of Medicine, 4 July 1985. Web. 27 July 2017.
16. Cullen, B. R. "Viruses and MicroRNAs." Nature Genetics. U.S. National Library of Medicine, June 2006. Web. 27 July 2017.
17. Denli, A. M., B. B. Tops, R. H. Plasterk, R. F. Ketting, and G. J. Hannon. "Processing of Primary MicroRNAs by the Microprocessor Complex." Nature. U.S. National Library of Medicine, 11 Nov. 2004. Web. 27 July 2017
18. Dollard SC, Wilson JL, Demeter LM, Bonnez W, Reichman RC, Broker TR, Chow LT. 1992. Production of human papillomavirus and modulation of the infectious program in epithelial raft cultures. OFF. Genes Dev **6**:1131–1142.
19. Dölken, L, et al. “Cytomegalovirus MicroRNAs Facilitate Persistent Virus Infection in Salivary Glands.” *PLoS Pathogens.*, U.S. National Library of Medicine, 14 Oct. 2010, www.ncbi.nlm.nih.gov/pubmed/20976200.
20. Elkon, R., A. P. Ugalde, and R. Agami. "Alternative Cleavage and Polyadenylation: Extent, Regulation and Function." Nature Reviews.

Genetics. U.S. National Library of Medicine, July 2013. Web. 27 July 2017.
21. Esquela-Kerscher, A., and F. J. Slack. "Oncomirs - MicroRNAs with a Role in Cancer." Nature Reviews. Cancer. U.S. National Library of Medicine, Apr. 2006. Web. 27 July 2017.
22. Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. "Potent and Specific Genetic Interference by Double-stranded RNA in Caenorhabditis Elegans." Nature. U.S. National Library of Medicine, 19 Feb. 1998. Web. 27 July 2017.
23. Fregonese, Laura, and Jan Stolk. "Hereditary Alpha-1-antitrypsin Deficiency and Its Clinical Consequences." Orphanet Journal of Rare Diseases. BioMed Central, 2008. Web. 27 July 2017.
24. Geoghegan, E. M., N. L. Welch, M. J. Yabsley, M. E. Church, P. A. Pesavento, and C. B. Buck. "Identification of a Second Raccoon-Associated Polyomavirus." *Genome Announcements.* U.S. National Library of Medicine, 29 June 2017. Web. 28 July 2017.
25. Gottwein, E., X. Cai, and B. R. Cullen. "A Novel Assay for Viral MicroRNA Function Identifies a Single Nucleotide Polymorphism That Affects Drosha Processing." Journal of Virology. U.S. National Library of Medicine, June 2006. Web. 27 July 2017.
26. Greco, D., N. Kivi, K. Qian, S. K. Leivonen, P. Auvinen, and E. Auvinen. "Human Papillomavirus 16 E5 Modulates the Expression of Host MicroRNAs." *PloS One.* U.S. National Library of Medicine, n.d. Web. 28 July 2017.
27. Greene, C. M., S. D. Miller, T. Carroll, C. McLean, M. O'Mahony, M. W. Lawless, S. J. O'Neill, C. C. Taggart, and N. G. McElvaney. "Alpha-1 Antitrypsin Deficiency: A Conformational Disease Associated with Lung and Liver Manifestations." Journal of Inherited Metabolic Disease. U.S. National Library of Medicine, Feb. 2008. Web. 27 July 2017.
28. Grishok, A., A. E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D. L. Baillie, A. Fire, G. Ruvkun, and C. C. Mello. "Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs That Control C. Elegans Developmental Timing." Cell. U.S. National Library of Medicine, 13 July 2001. Web. 27 July 2017.
29. Gu, Wenyi, Jiyuan An, Ping Ye, Kong-Nan Zhao, and Annika Antonsson. "Prediction of Conserved MicroRNAs from Skin and Mucosal Human Papillomaviruses." Archives of Virology 156.7 (2011): 1161-171. SpringerLink. Springer Vienna, 26 Mar. 2011. Web. 27 July 2017.

30. Gunasekharan, V., and L. A. Laimins. "Human Papillomaviruses Modulate MicroRNA 145 Expression to Directly Control Genome Amplification." *Journal of Virology.* U.S. National Library of Medicine, May 2013. Web. 28 July 2017.
31. Hassan, T., T. P. Carroll, P. G. Buckley, R. Cummins, S. J. O'Neill, N. G. McElvaney, and C. M. Greene. "MiR-199a-5p Silencing Regulates the Unfolded Protein Response in Chronic Obstructive Pulmonary Disease and α1-antitrypsin Deficiency." American Journal of Respiratory and Critical Care Medicine. U.S. National Library of Medicine, 01 Feb. 2014. Web. 28 July 2017.
32. Hassan, Tidi, Stephen G. J. Smith, Kevin Gaughan, Irene K. Oglesby, Shane O'Neill, Noel G. McElvaney, and Catherine M. Greene. "Isolation and Identification of Cell-specific MicroRNAs Targeting a Messenger RNA Using a Biotinylated Anti-sense Oligonucleotide Capture Affinity Technique." Nucleic Acids Research. Oxford University Press, Apr. 2013. Web. 28 July 2017.
33. Hutvágner, G., J. McLachlan, A. E. Pasquinelli, E. Bálint, T. Tuschl, and P. D. Zamore. "A Cellular Function for the RNA-interference Enzyme Dicer in the Maturation of the Let-7 Small Temporal RNA." Science (New York, N.Y.). U.S. National Library of Medicine, 03 Aug. 2001. Web. 27 July 2017.
34. Jens, M., and N. Rajewsky. "Competition between Target Sites of Regulators Shapes Post-transcriptional Gene Regulation." Nature Reviews. Genetics. U.S. National Library of Medicine, Feb. 2015. Web. 27 July 2017.
35. Khvorova, A., A. Reynolds, and S. D. Jayasena. "Functional SiRNAs and MiRNAs Exhibit Strand Bias." Cell. U.S. National Library of Medicine, 17 Oct. 2003. Web. 28 July 2017.
36. Kincaid, R. P., and C. S. Sullivan. "Virus-encoded MicroRNAs: An Overview and a Look to the Future." PLoS Pathogens. U.S. National Library of Medicine, Dec. 2012. Web. 27 July 2017.
37. Kincaid, R. P., J. M. Burke, and C. S. Sullivan. "RNA Virus MicroRNA That Mimics a B-cell OncomiR." Proceedings of the National Academy of Sciences of the United States of America. U.S. National Library of Medicine, 21 Feb. 2012. Web. 27 July 2017.
38. Kincaid, R. P., J. M. Burke, J. C. Cox, E. M. De, and C. S. Sullivan. "A Human Torque Teno Virus Encodes a MicroRNA That Inhibits Interferon

Signaling." PLoS Pathogens. U.S. National Library of Medicine, 19 Dec. 2013. Web. 27 July 2017.
39. Kincaid, R. P., Y. Chen, J. E. Cox, A. Rethwilm, and C. S. Sullivan. "Noncanonical MicroRNA (miRNA) Biogenesis Gives Rise to Retroviral Mimics of Lymphoproliferative and Immunosuppressive Host MiRNAs." MBio. U.S. National Library of Medicine, 08 Apr. 2014. Web. 27 July 2017.
40. Lagos-Quintana, M., R. Rauhut, W. Lendeckel, and T. Tuschl. "Identification of Novel Genes Coding for Small Expressed RNAs." Science (New York, N.Y.). U.S. National Library of Medicine, 26 Oct. 2001. Web. 28 July 2017.
41. Landthaler, M., A. Yalcin, and T. Tuschl. "The Human DiGeorge Syndrome Critical Region Gene 8 and Its D. Melanogaster Homolog Are Required for MiRNA Biogenesis." Current Biology : CB. U.S. National Library of Medicine, 14 Dec. 2004. Web. 27 July 2017.
42. Larsson, C. "Natural History and Life Expectancy in Severe Alpha1-antitrypsin Deficiency, Pi Z." Acta Medica Scandinavica. U.S. National Library of Medicine, 1978. Web. 27 July 2017.
43. Lee, R. C., R. L. Feinbaum, and V. Ambros. "The C. Elegans Heterochronic Gene Lin-4 Encodes Small RNAs with Antisense Complementarity to Lin-14." Cell. U.S. National Library of Medicine, 03 Dec. 1993. Web. 27 July 2017.
44. Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Rådmark, S. Kim, and V. N. Kim. "The Nuclear RNase III Drosha Initiates MicroRNA Processing." Nature. U.S. National Library of Medicine, 25 Sept. 2003. Web. 27 July 2017.
45. Lee, Y., M. Kim, J. Han, K. H. Yeom, S. Lee, S. H. Baek, and V. N. Kim. "MicroRNA Genes Are Transcribed by RNA Polymerase II." The EMBO Journal. U.S. National Library of Medicine, 13 Oct. 2004. Web. 27 July 2017.
46. Lin, Y. T., and C. S. Sullivan. "Expanding the Role of Drosha to the Regulation of Viral Gene Expression." *Proceedings of the National Academy of Sciences of the United States of America.* U.S. National Library of Medicine, 05 July 2011. Web. 28 July 2017.
47. Lin, Y. T., R. P. Kincaid, D. Arasappan, S. E. Dowd, S. P. Hunicke-Smith, and C. S. Sullivan. "Small RNA Profiling Reveals Antisense Transcription throughout the KSHV Genome and Novel Small RNAs." *RNA (New York, N.Y.).* U.S. National Library of Medicine, Aug. 2010. Web. 28 July 2017.

48. Luisetti, M., and N. Seersholm. "α1-Antitrypsin Deficiency · 1: Epidemiology of α1-antitrypsin Deficiency." Thorax. BMJ Group, Feb. 2004. Web. 27 July 2017.
49. Maslakova, A. A., M. V. Telkov, I. V. Orlovsky, and O. S. Sokolova. "Comparative Analysis of SERPINA1 Gene Expression in Tumor Cell Lines." Moscow University Biological Sciences Bulletin 70.3 (2015): 127-31. SpringerLink. Allerton Press, 23 Sept. 2015. Web. 28 July 2017.
50. McClure, L. V., R. P. Kincaid, J. M. Burke, A. Grundhoff, and C. S. Sullivan. "Comprehensive Mapping and Analysis of Kaposi's Sarcoma-associated Herpesvirus 3' UTRs Identify Differential Posttranscriptional Control of Gene Expression in Lytic versus Latent Infection." *Journal of Virology.* U.S. National Library of Medicine, Dec. 2013. Web.
51. McClure, L. V., Y. T. Lin, and C. S. Sullivan. "Detection of Viral MicroRNAs by Northern Blot Analysis." Methods in Molecular Biology (Clifton, N.J.). U.S. National Library of Medicine, n.d. Web. 27 July 2017.
52. Melar-New, M., and L. A. Laimins. "Human Papillomaviruses Modulate Expression of MicroRNA 203 upon Epithelial Differentiation to Control Levels of P63 Proteins." *Journal of Virology.* U.S. National Library of Medicine, May 2010. Web. 28 July 2017.
53. Mendell, J. T., and E. N. Olson. "MicroRNAs in Stress Signaling and Human Disease."Cell. U.S. National Library of Medicine, 16 Mar. 2012. Web. 28 July 2017.
54. Moszyńska, Adrianna, Magdalena Gebert, James F. Collawn, and Rafał Bartoszewski. "SNPs in MicroRNA Target Sites and Their Potential Role in Human Disease." Open Biology. The Royal Society, Apr. 2017. Web. 27 July 2017.
55. Murphy, E, et al. “Suppression of Immediate-Early Viral Gene Expression by Herpesvirus-Coded MicroRNAs: Implications for Latency.” *Proceedings of the National Academy of Sciences of the United States of America.*, U.S. National Library of Medicine, 8 Apr. 2008, www.ncbi.nlm.nih.gov/pubmed/18378902.
56. Prosperi, A., M. Chiari, M. Zanoni, L. Gallina, G. Casà, A. Scagliarini, and A. Lavazza. "Identification and Characterization of Fringilla Coelebs Papillomavirus 1 (FcPV1) in Free-living and Captive Birds in Italy." Journal of Wildlife Diseases. U.S. National Library of Medicine, July 2016. Web. 27 July 2017.
57. Qian, Kui, Tuuli Pietilä, Mikko Rönty, Frederic Michon, Mikko J. Frilander, Jarmo Ritari, Jussi Tarkkanen, Lars Paulín, Petri Auvinen, and Eeva

Auvinen. "Identification and Validation of Human Papillomavirus Encoded MicroRNAs." PLoS ONE. Public Library of Science, 30 July 2013. Web. 27 July 2017.

58. Rosenbloom, K. R., J. Armstrong, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, G. Hickey, A. S. Hinrichs, R. Hubley, D. Karolchik, K. Learned, B. T. Lee, C. H. Li, K. H. Miga, N. Nguyen, B. Paten, B. J. Raney, A. F. Smit, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn, and W. J. Kent. "The UCSC Genome Browser Database: 2015 Update." *Nucleic Acids Research.* U.S. National Library of Medicine, Jan. 2015. Web. 28 July 2017.
59. Salzman, D. W., and J. B. Weidhaas. "SNPing Cancer in the Bud: MicroRNA and MicroRNA-target Site Polymorphisms as Diagnostic and Prognostic Biomarkers in Cancer." Pharmacology & Therapeutics.U.S. National Library of Medicine, Jan. 2013. Web. 27 July 2017.
60. Schiffman, M, et al. “Carcinogenic Human Papillomavirus Infection.” *Nature Reviews. Disease Primers.*, U.S. National Library of Medicine, 1 Dec. 2016, www.ncbi.nlm.nih.gov/pubmed/27905473.
61. Schwarz, D. S., G. Hutvágner, T. Du, Z. Xu, N. Aronin, and P. D. Zamore. "Asymmetry in the Assembly of the RNAi Enzyme Complex." Cell. U.S. National Library of Medicine, 17 Oct. 2003. Web. 28 July 2017.
62. Seo, G. J., L. H. L. Fink, B. O'Hara, W. J. Atwood, and C. S. Sullivan. "Evolutionarily Conserved Function of a Viral MicroRNA." Journal of Virology. American Society for Microbiology (ASM), Oct. 2008. Web. 27 July 2017.
63. Skalsky, R. L., and B. R. Cullen. "Viruses, MicroRNAs, and Host Interactions." *Annual Review of Microbiology.* U.S. National Library of Medicine, 10 Apr. 2010. Web. 28 July 2017
64. Skalsky, R. L., S. A. Barr, A. J. Jeffery, T. Blair, R. Estep, and S. W. Wong. "Japanese Macaque Rhadinovirus Encodes a Viral MicroRNA Mimic of the MiR-17 Family." *Journal of Virology.* U.S. National Library of Medicine, 29 Sept. 2016. Web
65. Sood, P., A. Krek, M. Zavolan, G. Macino, and N. Rajewsky. "Cell-type-specific Signatures of MicroRNAs on Target MRNA Expression." Proceedings of the National Academy of Sciences of the United States of America. U.S. National Library of Medicine, 21 Feb. 2006. Web. 27 July 2017.

66. Stern-Ginossar, N, et al. “Analysis of Human Cytomegalovirus-Encoded MicroRNA Activity during Infection.” *Journal of Virology.*, U.S. National Library of Medicine, Oct. 2009, www.ncbi.nlm.nih.gov/pubmed/19656885.
67. Stoller, J K, and L S Aboussouan. “Alpha1-Antitrypsin Deficiency.” *Lancet (London, England).*, U.S. National Library of Medicine, 25 June 2005, www.ncbi.nlm.nih.gov/pubmed/15978931.
68. Sullivan, C S, et al. “SV40-Encoded MicroRNAs Regulate Viral Gene Expression and Reduce Susceptibility to Cytotoxic T Cells.” *Nature.*, U.S. National Library of Medicine, 2 June 2005, www.ncbi.nlm.nih.gov/pubmed/15931223.
69. Taggart, C, et al. “Oxidation of Either Methionine 351 or Methionine 358 in Alpha 1-Antitrypsin Causes Loss of Anti-Neutrophil Elastase Activity.” *The Journal of Biological Chemistry.*, U.S. National Library of Medicine, 1 Sept. 2000, www.ncbi.nlm.nih.gov/pubmed/10867014.
70. Tycowski, K. T., Y. E. Guo, N. Lee, W. N. Moss, T. K. Vallery, M. Xie, and J. A. Steitz. "Viral Noncoding RNAs: More Surprises." *Genes & Development.* U.S. National Library of Medicine, 15 Mar. 2015. Web. 28 July 2017.
71. Virtanen, E., T. Pietilä, P. Nieminen, K. Qian, and E. Auvinen. "Low Expression Levels of Putative HPV Encoded MicroRNAs in Cervical Samples." SpringerPlus. U.S. National Library of Medicine, 22 Oct. 2016. Web. 27 July 2017.
72. Walboomers, J. M., M. V. Jacobs, M. M. Manos, F. X. Bosch, J. A. Kummer, K. V. Shah, P. J. Snijders, J. Peto, C. J. Meijer, and N. Muñoz. "Human Papillomavirus Is a Necessary Cause of Invasive Cervical Cancer Worldwide." *The Journal of Pathology.* U.S. National Library of Medicine, Sept. 1999. Web.
73. Wang, X., H. K. Wang, Y. Li, M. Hafner, N. S. Banerjee, S. Tang, D. Briskin, C. Meyers, L. T. Chow, X. Xie, T. Tuschl, and Z. M. Zheng. "MicroRNAs Are Biomarkers of Oncogenic Human Papillomavirus Infections." Proceedings of the National Academy of Sciences of the United States of America. U.S. National Library of Medicine, 18 Mar. 2014. Web. 27 July 2017.
74. Weber, K., M. E. Bolander, and G. Sarkar. "PIG-B: A Homemade Monophasic Cocktail for the Extraction of RNA." *Molecular Biotechnology.* U.S. National Library of Medicine, Feb. 1998. Web.
75. Yi, R., Y. Qin, I. G. Macara, and B. R. Cullen. "Exportin-5 Mediates the Nuclear Export of Pre-microRNAs and Short Hairpin RNAs." Genes &

Development. U.S. National Library of Medicine, 15 Dec. 2003. Web. 27 July 2017.
76. You, L., J. Wu, Y. Feng, Y. Fu, Y. Guo, L. Long, H. Zhang, Y. Luan, P. Tian, L. Chen, G. Huang, S. Huang, Y. Li, J. Li, C. Chen, Y. Zhang, S. Chen, and A. Xu. "APASdb: A Database Describing Alternative Poly(A) Sites and Selection of Heterogeneous Cleavage Sites Downstream of Poly(A) Signals." *Nucleic Acids Research.* U.S. National Library of Medicine, Jan. 2015. Web. 28 July 2017.
77. Zeng, Y., R. Yi, and B. R. Cullen. "Recognition and Cleavage of Primary MicroRNA Precursors by the Nuclear Processing Enzyme Drosha." The EMBO Journal. U.S. National Library of Medicine, 12 Jan. 2005. Web. 27 July 2017.
78. Zhao, Wenxue, Denitza Blagev, Joshua L. Pollack, and David J. Erle. "Toward a Systematic Understanding of MRNA 3′ Untranslated Regions." Proceedings of the American Thoracic Society. American Thoracic Society, 01 May 2011. Web. 27 July 2017.
79. Zheng, Z. M., and X. Wang. "Regulation of Cellular MiRNA Expression by Human Papillomaviruses." *Biochimica Et Biophysica Acta.* U.S. National Library of Medicine, Nov. 2011. Web. 28 July 2017.
80. Zisoulis, D G, et al. “Comprehensive Discovery of Endogenous Argonaute Binding Sites in Caenorhabditis Elegans.” *Nature Structural & Molecular Biology.*, U.S. National Library of Medicine, Feb. 2010, www.ncbi.nlm.nih.gov/pubmed/20062054.
81. Zur, H. "Papillomaviruses and Cancer: From Basic Studies to Clinical Application." *Nature Reviews. Cancer.* U.S. National Library of Medicine, May 2002. Web. 28 July 2017.