

Identification of Human MicroRNA-Like Sequences Embedded within the Protein-Encoding Genes of the Human Immunodeficiency Virus

Bryan Holland, Jonathan Wong, Meng Li, Suraiya Rasheed*

Laboratory of Viral Oncology and Proteomics Research, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America

Abstract

Background: MicroRNAs (miRNAs) are highly conserved, short (18–22 nts), non-coding RNA molecules that regulate gene expression by binding to the 3′ untranslated regions (3′UTRs) of mRNAs. While numerous cellular microRNAs have been associated with the progression of various diseases including cancer, miRNAs associated with retroviruses have not been well characterized. Herein we report identification of microRNA-like sequences in *coding* regions of several HIV-1 genomes.

Results: Based on our earlier proteomics and bioinformatics studies, we have identified 8 cellular miRNAs that are predicted to bind to the mRNAs of multiple proteins that are dysregulated during HIV-infection of CD4+ T-cells *in vitro*. *In silico* analysis of the full length and mature sequences of these 8 miRNAs and comparisons with all the genomic and subgenomic sequences of HIV-1 strains in global databases revealed that the first 18/18 sequences of the *mature* hsa-miR-195 sequence (including the short *seed* sequence), matched perfectly (100%), or with one nucleotide mismatch, *within* the envelope (*env*) genes of five HIV-1 genomes from Africa. In addition, we have identified 4 other miRNA-like sequences (hsa-miR-30d, hsa-miR-30e, hsa-miR-374a and hsa-miR-424) within the *env* and the *gag-pol* encoding regions of several HIV-1 strains, albeit with reduced homology. Mapping of the miRNA-homologues of *env* within HIV-1 genomes localized these sequence to the functionally significant variable regions of the *env* glycoprotein gp120 designated V1, V2, V4 and V5.

Conclusions: We conclude that microRNA-like sequences are embedded within the protein-encoding regions of several HIV-1 genomes. Given that the V1 to V5 regions of HIV-1 envelopes contain specific, well-characterized domains that are critical for immune responses, virus neutralization and disease progression, we propose that the newly discovered miRNA-like sequences within the HIV-1 genomes may have evolved to self-regulate survival of the virus in the host by evading innate immune responses and therefore influencing persistence, replication and/or pathogenicity.

Citation: Holland B, Wong J, Li M, Rasheed S (2013) Identification of Human MicroRNA-Like Sequences Embedded within the Protein-Encoding Genes of the Human Immunodeficiency Virus. PLoS ONE 8(3): e58586. doi:10.1371/journal.pone.0058586

Editor: Fatah Kashanchi, George Mason University, United States of America

Received: August 22, 2012; **Accepted:** February 5, 2013; **Published:** March 8, 2013

Copyright: © 2013 Holland et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by the University of Southern California-Rasheed Research Endowment Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: srasheed@usc.edu

Introduction

MicroRNAs (miRNAs) are highly conserved, naturally occurring, 18–22 nucleotides long, noncoding RNA molecules that are processed from long precursor transcripts (pre-miRNAs). The cellular miRNAs are transcribed by the same transcription factors and RNA polymerase II that control the transcription of protein-encoding messenger RNAs (mRNAs) [1]. The pre-miRNA transcripts are processed into mature miRNAs by complex molecular processes involving cleavage by RNaseIII-like enzyme Drosha in the nucleus, followed by their transport to the cytoplasm mediated by Exportin-5 and a second cleavage by Dicer, an RNase III-like enzyme. One strand of this double-stranded miRNA forms the “mature” 18–22 nucleotide sequence, which becomes a part of a multi-protein RNA-induced silencing-complex (RISC) that guides it to the mRNA target site. The second strand is believed to be degraded [1].

The human genome contains >1,500 miRNA genes that have been predicted or experimentally shown to play critical roles in

normal cellular functions, such as maintaining homeostasis, and regulating or modulating viral and cellular gene expression. Specific alterations in gene or protein expression profiles change the direction of normal cells toward the development of cancer and other disorders [2,3]. An important post-transcriptional regulatory step in gene expression is that the 5′ ends of miRNAs can base-pair with the complementary sequences in the 3′ untranslated regions (UTRs) of their target mRNAs and suppress translational capacities of those mRNAs [4]. While the exact mechanism of miRNA-mediated regulation of the mRNA targets is still not fully understood, two aspects of the gene-expression controls are noteworthy. First, a short nucleotide sequence (nucleotides 2–7 from the 5′ end of a miRNA), referred to as the ‘*seed sequence*’, base-pairs completely and continuously with its target mRNA; next, this initial binding of the miRNA with its *seed sequence* allows complete or incomplete binding of the rest of the mature miRNA sequence with the target mRNA. Thus, gene repression by miRNA appears to depend primarily on complete

base-pairing of their *seed sequences* with its complementary mRNA target; however, sometimes an incomplete base-pairing of the rest of the miRNA sequence also destabilizes the target transcript, represses the translation of the protein or influences degradation of the target mRNA [5]. Also, one miRNA can interact with multiple mRNA targets and a single mRNA can be regulated by multiple miRNAs. Upregulation or downregulation of a given miRNA can dysregulate protein expression profiles and therefore result in disruption of normal biological processes such as cell proliferation, development, differentiation, apoptosis, and signal transduction [4,6,7].

The bulk of well-characterized miRNAs are either from cellular genomes or from several DNA *viral genomes* and both can impact expression of cellular and viral genes respectively in infected cells [8]. Recently, a retrovirus, the bovine leukemia virus (BLV), has been shown to encode a conserved cluster of miRNAs that are transcribed by RNA polymerase III which mimics miRNAs involved in B-cell cancers [9].

In silico studies have predicted that the genomes of the human immunodeficiency virus-1 (HIV-1) may contain *target sites* for binding of *cellular miRNAs* [10]. For instance, hsa-miR-29a is reported to *target the nef region* of the HIV-1 genome [11]. Researchers have also shown that cellular miRNA expression patterns in HIV-1 infected peripheral blood mononuclear cells are altered post-infection [12,13]. A recent report proposed that altered cellular miRNA profiles of HIV-infected cells could be used as early indicators of host cellular dysfunctions [14]. The existence of HIV-1 vmiRNAs, that may regulate both viral and host gene expression has also been suggested [15].

Based on the RNA structure and folding parameters consistent with cellular miRNA molecules, *five HIV-1-encoded miRNAs have been predicted* computationally [16]. However, there has been disagreement about the existence of vmiRNAs in HIV-1 genomes [17]. Currently, three HIV-1-encoded vmiRNAs have been listed in miRBase: miR-H1, miR-N367 and miR-TAR-5p/3p. These vmiRNAs have been reported to *target "viral" transcripts*, as in the case of miR-N367 [18], and host cellular factors for miR-H1 and miR-TAR-5p/3p [19,20]. More recent research seems to support that miRNAs are in fact present in the viral genomes: Ouellet et al. [21] showed that functional miRNAs are processed from the HIV TAR element; Yeung et al. [22] have identified small non-coding RNAs in HIV infected cells and corroborated the existence of virally encoded miRNAs in *nef* and *TAR*; and Schopman et al. [23] have identified HIV-encoded small RNAs in virus-infected cells. While studies have reported that human endogenous retroviruses, retroelements and several exogenous retroviral sequences may have homologies with cellular miRNAs [24], the viral homologues of *human miRNAs in HIV-1 genomes* have not been identified and the role of miRNAs in HIV-1 infection has not been elucidated.

Our earlier proteomics and bioinformatics studies had identified a significant number of functionally relevant proteins that were upregulated, downregulated or produced *de novo* in a chronically HIV-infected CD4+ clonal T-cell line (RH9) [25,26]. Based on these findings we hypothesized that the differential expression of proteins in HIV-infected versus uninfected counterpart cells is due to the dysregulation of cellular (or viral) microRNAs. Using multiple bioinformatics and computational tools we have identified 8 miRNAs that are predicted to bind to the mRNA targets of several proteins that are differentially expressed in HIV-infected T-cells (Table 1). To understand the biological significance of these miRNAs we searched for nucleotide sequence similarities between each of the miRNAs identified, and the whole or partial HIV-1 genome sequences present in the global databases. Herein,

we report the discovery of a *sequence homologue* of hsa-miR-195 *within the HIV-1 protein-coding regions* of several HIV-1 strains from Africa (Table 2). In addition, we have identified sequence similarities between 3 other mature human miRNAs (hsa-miR-30d, hsa-miR-374a and hsa-miR-424) (Table 3) and one full-length miRNA (hsa-miR-30e) within the HIV-1 envelope regions (Table 4). A detailed search of the whole HIV-1 genome sequence also displayed similarity between 2 full-length human miRNAs (hsa-miR-30d and hsa-miR-424) and the gag-pol protein-encoding regions or pol only of 4 HIV-1 isolates (Table 4). These data have provided new insights into possible roles of viral env-gene associated miRNA-like sequences in the survival of HIV-1 genomes in the host, immune responses, virus replication, and pathogenesis.

Results

Identification of Cellular microRNA Sequences Associated with Dysregulated Proteins in HIV-Infected Cells

We used the program "GeneSet2miRNA" to identify potential microRNAs that could impact the activities of the proteins whose expression had been shown to be dysregulated in HIV-infected RH9 T-cells [25,26]. For example, the microtubule-actin cross-linking factor 1 (MACF1) [27] is one of the proteins which we found was downregulated due to HIV-infection. Bioinformatics analyses indicated that the 3' untranslated region of MACF1 has a sequence (5' GGACAAUAGCUGCUA 3') which is complementary to the seed sequence for the mature hsa-miR-195 (5' UAGCAGCACAGAAU 3'), indicating a strong binding affinity to this protein. MACF1 has an actin-regulated ATPase activity necessary for cross-linking actin to other cytoskeletal proteins. In addition, MACF isoforms act as positive regulators of Wnt receptor signaling pathway and are essential for controlling focal adhesions assembly. Actin-related proteins also play critical roles during HIV infection and latency in T-cells ([28] and our unpublished data).

Using an adjusted p-value of the enrichment (adjusted for multiple testing by Monte-Carlo simulations) cutoff of 0.05, we identified 7 microRNAs that could significantly bind to multiple mRNA targets in these cells. In addition, we selected one more miRNA from our initial search which fulfilled the criteria of being the best single-model match of the GeneSet2miRNA program. These 8 miRNA sequences are listed in Table 1. These analyses also revealed that hsa-miR-195 was one of the most frequently detected cellular miRNAs among the 8 miRNAs identified to be associated with mRNAs of the proteins modulated in our experimentally HIV-1 infected T-cells. Each of the miRNA and mRNA interactions were highly significant among 7 of the 8 miRNAs identified (p-value = 0.002 to 0.008).

Discovery of miRNA-195 Homologues in the HIV-ENV Gene

Using both the mature (~22 nucleotides) and full length (~70 nucleotides) sequences of each of the 8 cellular miRNAs listed in Table 1 as query sequences, we have used multiple bioinformatics tools including BLAST (Basic Local Alignment Search Tool) and Clustal to align and scrutinize potential matches in HIV-1 genomes. An extensive search of more than 3000 complete HIV-1 genomes and over 400,000 HIV-1 subgenomic sequences present in both the Los Alamos and NCBI databases resulted in thousands of potential matches. Each one of these matches was subsequently screened on the basis of the degree of homology and degree of sequence continuity with the full length and mature sequences of each of the 8 previously identified cellular miRNAs. These analyses resulted in the discovery of a perfect match (100%)

Table 1. MicroRNA Sequences Used For Analysis.

microRNA	Length of miRNA : Full Length/Mature	miRBase Accession # : Full Length/Mature	Chromo-somal Location	Mature Cellular miRNA Sequence
hsa-miR-195	87 nt	MI0000489	17p13.1	UAGCAGCACAGAAAUAUUGGC
	21 nt	MIMAT0000461		
hsa-miR-15b	98 nt	MI0000438	3q25.33	UAGCAGCACAUCAUGGUUACA
	22 nt	MIMAT0000417		
hsa-miR-16-1	89 nt	MI0000070	13q14.2	UAGCAGCACGUAUUUUGGCG
	22 nt	MIMAT0000069		
hsa-miR-16-2	81 nt	MI0000115	3q25.33	UAGCAGCACGUAUUUUGGCG
	22 nt	MIMAT0000069		
hsa-miR-30d	70 nt	MI0000255	8q24.22	UGUAAACAUCCTCCGACUGGAAG
	22 nt	MIMAT0000245		
hsa-miR-30e	92 nt	MI0000749	1p34.2	UGUAAACAUCUUGACUGGAAG
	22 nt	MIMAT0000692		
hsa-miR-374a	72 nt	MI0000782	Xq13.2	UUUAUAUACAACCGUAAGUG
	22 nt	MIMAT0000727		
hsa-miR-424	98 nt	MI0001446	Xq26.3	CAGCAGCAAUUCUGUUUUGAA
	21 nt	MIMAT0001341		

doi:10.1371/journal.pone.0058586.t001

from nucleotide positions 1 through 18 of the mature hsa-miR-195 sequence, including the seed region at nucleotide positions 2–7, within the envelope (*env*) protein-coding sequences of an HIV-1 isolate from South Africa (accession #GU216763) (Table 2). Two other isolates, #GU216768 and #GU216773, from the same individual also showed homology with hsa-miRNA-195 in the same region of the Env gene and again matched miR-195 at positions 1–18, with 17 of 18 (94%) nucleotides matching perfectly (Table 2). This single nucleotide mutation (i.e. 17 versus 18

nucleotides) represents the genetic diversity due to mutations known to be present in *all* HIV-1 envelope genes of both acute and chronically HIV-1 infected individuals [29]. In addition to the sequence homology in the South African HIV-1 isolates (#GU216763), we have identified two other strains from Tanzania, #HM215313 and #DQ199139, whose Env gene sequences also matched with the same region of hsa-miR-195 at 17 of 18 nucleotide positions, as in the GU216763 isolate (Table 2).

Table 2. Sequence Similarity Between Human miR-195 and HIV Envelope Genes.

Gene Name/Description	Accession #/ Subtype/Country of Origin	Alignment - Top line is hsa-miR195 (MI0000489) mature sequence, and bottom line is the env region of the respective virus.	Number and Percentage of Nucleotide Matches	Seed Matches
HIV-1 Isolate	GU216763	1 tagcagcacagaaatatt 18	18/18	6/6
#169b1a12	Subtype C		(100%)	100%
Envelope gene	South Africa	1401 tagcagcacagaaatatt 1418		
HIV-1 Isolate	GU216768	1 tagcagcacagaaatatt 18	17/18	5/6
#169b1c4	Subtype C		(94%)	83%
Envelope gene	South Africa	1404 tagcggcacagaaatatt 1421		
HIV-1 Isolate	GU216773	1 tagcagcacagaaatatt 18	17/18	5/6
#169b1d7	Subtype C		(94%)	83%
Envelope gene	South Africa	1392 tagcggcacagaaatatt 1409		
HIV-1 Isolate	HM215313	1 tagcagcacagaaatatt 18	17/18	5/6
#401-F1_8_10	Subtype CD		94%	83%
Envelope gene	Tanzania	1362 tagcaccacagaaatatt 137		
HIV-1 Isolate	DQ199139	1 tagcagcacagaaatatt 18	17/18	6/6
#TZB0573	Subtype C		94%	100%
Envelope gene	Tanzania	585 tagcagcacagaaacatt 602		

HIV-1 isolate #169b1a12, #169b1c4, and #169b1d7 are from the same patient.

doi:10.1371/journal.pone.0058586.t002

Table 3. Sequence Similarity Between Three Mature Human microRNAs and HIV Envelope Genes.

Cellular miRNA (Mature Sequence)	Accession #/ Subtype/Country of Origin	Gene Name/ Description	Alignment - Top line is hsa-miR sequence, and bottom line is the env region of the respective virus.	Number and Percentage of Nucleotide Matches	Seed Matches
hsa-miR-30d	AY169802	HIV-1 Strain	3 taaacatccccga 15	13/13	5/6
	Group O	#98CMA104		(100%)	(83%)
	Cameroon	Complete genome	6901 taaacatccccga 688		
hsa-miR-374a	AJ429907	HIV-1 Strain	4 taatacaacctgataag 20	13/17	4/6
	Group M	#00NE079		(76%)	(67%)
	Subtype 6cpx	Envelope	175 taataccaattgatcag 191		
	Niger				
	AF391235	HIV-1 Clone	2 tataatacaacctgataa 19	17/18	6/6
	Group M	#TV006c9.1		(94%)	(100%)
hsa-miR-424	GU080167	HIV-1 Clone	6 gcaattcatgtttt 19	14/14	2/6
	Group M	704MC009F		(100%)	(33%)
	Subtype C	Envelope	417 gcaattcatgtttt 404		
	South Africa				

doi:10.1371/journal.pone.0058586.t003

Identification of HIV-ENV Gene Homology Domains in Other Cellular microRNAs

Analyses of the remaining seven cellular miRNAs that were predicted to be associated with multiple dysregulated proteins expressed in our experimentally HIV-infected cells (Table 1) also showed sequence similarities between 4 microRNA sequences and the *env* regions of several other HIV-1 strains (Tables 3, 4). Each of the seven miRNAs were screened using the BLAST algorithm to find the best matches within all the complete and partial HIV-1 genome sequences contained in both NCBI and the Los Alamos HIV databases. Again, several thousand preliminary matches were screened for the degree of homology, degree of continuity and presence of the seed sequence. These analyses indicated that in addition to hsa-miR-195, there are three other cellular mature miRNAs (hsa-miR-424, hsa-miR-374a, hsa-miR-30d) that displayed 13–18 nucleotide length homologous regions within the Env genes of four different HIV-1 strains (Table 3). We have also tested full-length (~70 nucleotides) miRNAs and found that hsa-miR-30e exhibited sequence homology within the *env* regions of 5 different HIV-1 isolates (Table 4). These sequences were distinct from those aligned with miR-195 and were located in different regions of the HIV-1 envelope than hsa-miR-195. Thus, the number, length, and percentage of matches within the seed region of each miRNA and their respective HIV *env*-gene sequence were variable.

Identification of miRNA-Homology Domains in other Regions of HIV-genomes

To determine if any other region of the HIV-1 genome might show homology domains, all the preliminary alignments were screened again using the query sequences of each of the 8 cellular miRNAs that we had identified earlier. Our results indicated that 2 out of the 8 predicted miRNAs (hsa-miR-424 and hsa-miR-30d), showed homologies in different regions within the HIV-1 genome (Table 4). Both full-length cellular miRNAs (hsa-miR-424 and hsa-

miR-30d) exhibited 76%–100% homology domains within the gag-pol regions of four isolates from three HIV-1 strains (Table 4). These findings suggest that multiple cellular miRNAs have homology domains in different regions of various HIV-1 genomes (Tables 2, 3 & 4). Thus, out of the 8 miRNAs that we identified for further analysis, 5 (hsa-miR-195, hsa-miR-424, hsa-miR-30d, hsa-miR-30e and hsa-miR-374a) showed homology domains in the HIV-1 genomes of several strains, and 3 (hsa-miR-15b, hsa-miR-16-1 and hsa-miR-16-2) did not.

Clustal Analyses and Mapping of MiRNA-like Sequences

Since the most significant matches in our study occurred between the cellular miRNA hsa-miR-195 and the *env* protein encoding regions of HIV-1 isolates from Africa, we further scrutinized and characterized these sequences by aligning the genomes of different HIV-1 clades from various regions around the world with the hsa-miR-195-like sequence homology domains. Using the ClustalW2 algorithm, as well as the Los Alamos compendium of all HIV-1 alignments, we were able to show the exact regions of these sequences in each of the 6 clades that corresponded to the position of the hsa-miR-195-like sequence we have identified from the African HIV-1 strains (Table 5).

We chose 17 HIV-1 genomes that had been identified as *representative strains* from each of the 6 HIV-1 clades (A, B, C, D, E and G) to perform alignments, and used whole genome sequences to produce the best alignment. These analyses revealed that the hsa-miR-195-like region that we have identified aligns best to a specific region in the Env gene of African strains (Table 5, shaded area). However, while miR-195-like sequences were shared by all representative strains among the 6 clades that we examined, these sequences were highly divergent among the different clades (Table 5).

In addition to defining the specificity of sequence alignments, we used the TreeDyn software for the construction of a sequence-based relational tree using the alignment data generated by the

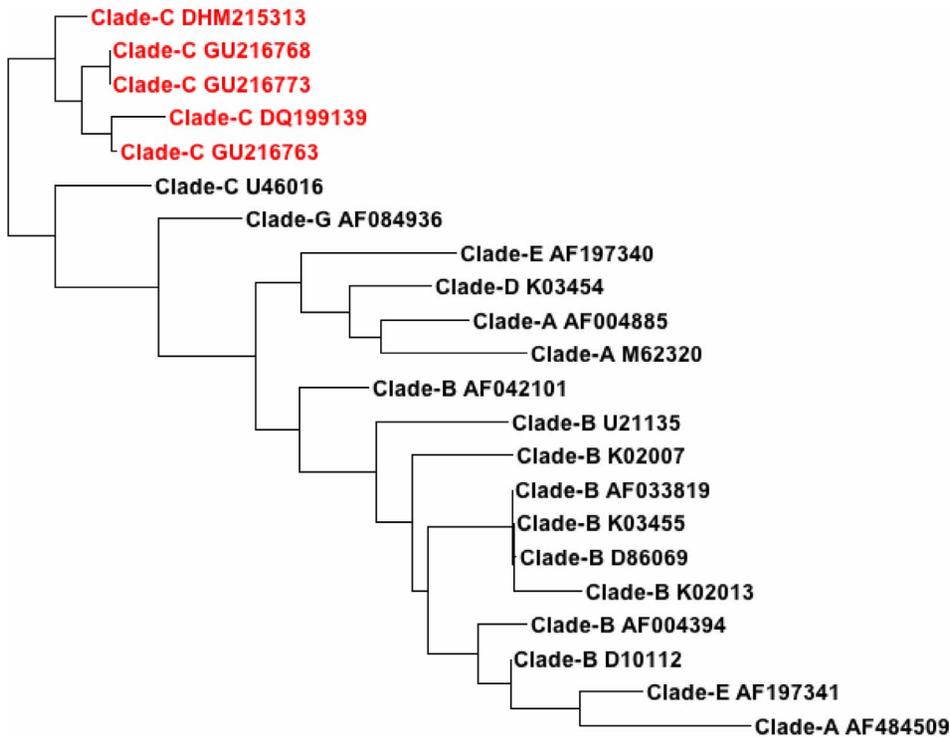


Figure 1. Phylogenetic Tree of miR-195-like sequences in different HIV clades. This figure shows the envelope region of the HIV HXB2 genome to which the hsa-miR-195-like sequence maps. The area corresponding to the V5 region of HXB2 spans nucleotide positions 7603 through 7632 and is highlighted in yellow. The hsa-miR-195-like region is shown in blue. The mismatches between the HXB2 sequence and the miR-like sequence present in the African strain #GU216763 are due to the sequence variations in envelope regions of all HIV genomes. doi:10.1371/journal.pone.0058586.g001

The miRNA-like sequences we have identified in HIV-1 are unique in that they do *not* seem to be derived from cellular miRNA, nor do they appear to represent viral miRNAs or their targets. The hsa-miR-195-like sequence corresponds to the first 18 nucleotides of the mature hsa-miR-195, which has a length of 21 nucleotides. Any functional similarity of this sequence to the cellular hsa-miR-195 may be speculative. However, it should be noted that there are miRNAs as short as 17 nt; and that HIV-1 has been reported to encode a viral miRNA, designated TAR-3p, whose cloned length is 17 nt [20]. Further, the potential action of a miRNA is mostly dependent on base pairing between the miRNA seed sequence and its target; positions 13–16 of the miRNA may aid in pairing as well [5,36]. The miR-195-like sequence we have identified in #GU216763 contains both the seed region and positions 13–16 of hsa-miR-195 and is 100% conserved in these regions. It has also been predicted computationally that the cellular hsa-miR-195 may interact with the HIV-1 Nef in the 3' LTR region based on a perfect complementarity of a 7 nucleotide seed sequence with its viral target [31].

Whether the microRNA-like sequences are of viral origin or products of provirus integration millions of years ago remains to be explored. However, the fact that this sequence is part of a functional coding region of a vital viral gene makes the integration event scenario seem unlikely. Our data suggest that the miR-like sequences present in the HIV-1 envelope region are *viral RNA sequences, which emulate a cellular miRNA*.

The implications of a viral miRNA mimicking a cellular miRNA would be speculative. However, there is precedence for a viral miRNA to outcompete its cellular competitor. HIV-1 TAR RNA has been reported to act as a sort of miRNA-‘decoy’, decreasing the host cell’s RNAi activity by binding and

sequestering TRBP, a TAR RNA-binding protein and an essential Dicer-cofactor [37].

We therefore propose that the hsa-miR-like sequences we have identified in the Env genes of several viruses may similarly be titrating out the related cellular miRNA targets.

A major point of distinction between miRNAs and our newly identified miRNA-like sequences is that while cellular miRNAs are derived from *non-coding* regions of the DNA, the miRNA-like sequences we have identified are located in the *coding* regions of vital HIV-1 genes. The HIV-1 envelope glycoprotein contains 5 variable regions (V1–V5) interspersed by conserved regions C1–C5. The miRNA-like sequences we discovered have been mapped to the V1, V2, V4 and V5 regions of the HIV-1 envelope and are integral components of the HIV-1 gene, which codes for a functional envelope gp120 (Figure 3). The finding of several viral sequences homologous to cellular miR-30d, miR-30e, miR-374a, miR-424 and miRNA-195 in different regions of the HIV-1 genome but primarily in the envelope regions of several HIV-1 strains indicate that the phenomenon of cellular miRNA-like sequences in the HIV-1 genome may be widespread.

Changes in the length of amino acid sequences or glycosylation patterns in the variable V-regions are critical to HIV-1 infection because they can affect not only the cellular tropism but can also modulate sensitivity to virus neutralization and disease progression [38]. Some of the major determinants that contribute to biological activities of HIV-1 strains including replication, viral tropism (ability to infect T-cells versus macrophages or other cell types), sensitivity to neutralization, modulation of the CD4 antigen, and cytopathogenicity are localized in the V1 to V5 regions of the HIV-1 envelope glycoprotein gp120 [39–41]. These sequences are critical for effective humoral responses and virus neutralization.

HXB2 Nucleotide Position	HXB2 env K03455 Nucleotide	V5 Region of HXB2	Cellular hsa-miR-195 Sequence (Position 1-18)	miR-195-like Sequence in HIV env
7600	G			
7601	G			
7602	T			
7603	A	A		
7604	A	A		
7605	T	T		
7606	A	A		
7607	G	G		
7608	C	C		
7609	A	A		
7610	A	A		
7611	C	C	T	T
7612	A	A	A	A
7613	A	A	G	G
7614	T	T	C	C
7615	G	G	A	A
7616	A	A	G	G
7617	G	G	C	C
7618	T	T	A	A
7619	C	C	C	C
7620	C	C	A	A
7621	G	G	G	G
7622	A	A	A	A
7623	G	G	A	A
7624	A	A	A	A
7625	T	T	T	T
7626	C	C	A	A
7627	T	T	T	T
7628	T	T	T	T
7629	C	C		
7630	A	A		
7631	G	G		
7632	A	A		
7633	C			
7634	C			
7635	T			

Figure 2. Location of hsa-miR-195-like Sequence in HXB2 Env Gene. Phylogenetic tree was constructed using the hsa-miR-195 sequence TAGCAGCACAGAAATATT for alignment similar to that used in Table 5. This tree was constructed using the treedyn program. Each leaf represents one of the 5 African microRNA-like sequences or one of the 17 reference strains from the 6 clades which were used for alignment. The five sequences highlighted in red represent the 5 African microRNA-like sequences identified in Table 2. doi:10.1371/journal.pone.0058586.g002

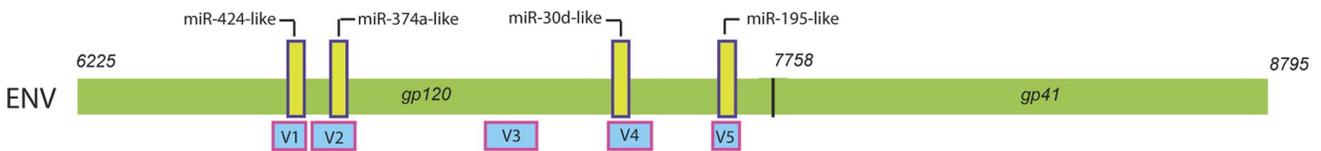


Figure 3. Localization of miR-like sequences in HIV envelope variable regions. This figure shows the miR-like domains listed in Tables 2 and 3 and mapped to their corresponding positions in the HXB2 envelope gene (at nucleotide positions 6682–6694, 6765–6782, 7386–7398 and 7611–7628). The miR-424-like, miR-374a-like, miR-30d-like, and miR-195-like sequences are embedded in the hypervariable regions corresponding to V1, V2, V4 and V5 respectively and are depicted in blue. No miRNA-like sequence was detected in the V3 region of the envelope. doi:10.1371/journal.pone.0058586.g003

Table 6. MicroRNA-like sequence alignments with cellular genome sequences.

miR-195-like:	1 TAGCAGCACAGAAATATT 18
Human chr 17:	6524380 TAGCAGCACAGAAATATT 6524363
miR-30d-like:	1 TAAACATCCCGA 13
Human chr 8:	49090730 TAAACATCCCGA 49090718
miR-374a-like:	1 TATAATACAACCTGATAA 18
Human chr X:	11825168 TATAATACAACCTGATAA 11825151
miR-424-like:	1 GCAATTCATGTTTT 14
Human chr X:	17948436 GCAATTCATGTTTT 17948423

BLAST search result and alignment using miR-195-like sequence against the human genome sequence and locations on the chromosomal DNAs.
doi:10.1371/journal.pone.0058586.t006

The V1 to V5 regions of the HIV-1 envelope have been associated with the rate of replication, virus neutralization and pathogenesis of HIV-1 strains [42,43]. While the primary virus replicates in the body, it becomes resistant to neutralization because some of the specific V1 to V5 domains are either lost or have been modified by increased numbers of N-linked glycosylation sites and therefore cellular immune responses to HIV-1 infection are compromised [39–41].

Finally, our findings of four different miRNA-like sequences within the V1, V2, V4 and V5 regions of the HIV-1 *env* gene may provide new insights that will contribute to a better understanding of the molecular complexities of HIV-1 infection and pathogenesis. The miRNA-like sequences may emulate a non-coding cellular miRNA and therefore could represent the first examples of *human cellular homologues* of miRNAs in HIV-1 coding regions. These sequences may therefore play a role in HIV-1 replication, immunity, and virus neutralization, and thus may influence pathogenesis in HIV-infected individuals. Detailed *in vivo* studies and construction of vectors containing the human microRNA-like sequences that we have discovered, would yield critical results that would allow us to develop a humanized mouse model system to test the effects of these vectors *in vivo*. These experiments will demonstrate the influence of microRNA-like sequences on virus replication, as well as antibody and antigen production *in vitro* and *in vivo*.

Materials and Methods

Source of Data

Previous proteomics and bioinformatics research in our laboratory had identified >200 differentially expressed, functionally relevant proteins in an HIV-1 infected CD4+ T-cell line (RH9) analyzed sequentially over a period of approximately 2 years [25,26]. In this study, we used GeneSet2miRNA [44] to identify potential microRNAs that could impact the activities of our differentially regulated proteins. Using an adjusted p-value of the enrichment (adjusted for multiple testing by Monte-Carlo simulations) cutoff of 0.05, we identified 7 miRNAs that may significantly bind to multiple mRNA targets. We also selected one

other miRNA because it was identified as the best single-model match. These 8 miRNAs are listed in Table 1.

Identification of Homologous Sequences

To identify sequences that could be homologous to HIV-1, we downloaded the *full length and mature sequences* of the 8 human microRNAs from the miRBase database (<http://mirbase.org/>) that had been shown to be significantly associated with the proteins modulated by HIV-infection of CD4+ T-cells. Each of the 8 miRNAs was used as a separate query, utilizing both the mature and full-length versions of each miRNA. The BLAST (Basic Local Alignment Search Tool) [45] program was used to search against the entire HIV-1 databases (<http://blast.ncbi.nlm.nih.gov/>) (HIV taxid: 11676) at the National Center for Biotechnology Information (NCBI) and the Los Alamos HIV databases (http://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html). All full-length and partial HIV-1 genome sequences, representative of all HIV-1 clades and strains, were used for the analyses. These sequences have been identified by the International Committee on Taxonomy of viruses (ICTV) and are available in the global public databases. The outputs from both database searches were compared and the best matches from all microRNA query searches were selected based on the length of the match, percentage of identity of match, lack of gaps or deletions, and inclusion of the seed sequence.

Clustal Analyses and Mapping of Newly Identified Sequences

The Clustal algorithm was used for multiple sequence alignments [46,47] (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). We used the five most homologous HIV-1 sequences to hsa-miR-195 as identified by our BLAST searches and shown in Table 2, as well as the sequences of 17 other representative HIV-1 strains from 6 clades to generate alignments using the Clustal algorithm. Complete genome sequences from each of the representative HIV-1 strains were used to perform alignments of the different clades with our best matches. Results from the Clustal algorithm were then checked against the Los Alamos HIV Compendium (<http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/compendium.html>) to verify that the alignments from both sources were in agreement.

In addition to defining the specificity of sequence alignments, we used the TreeDyn software program for the construction of a sequence-based relational tree using the alignment data generated by the Clustal algorithm (<http://www.treedyn.org/>). The target regions of the alignment were then mapped to the HXB2 strain gene map using the Los Alamos National Laboratory HIV genome database (<http://www.hiv.lanl.gov/>) map, because this is one of the most complete reference sequence data maps available for HIV-1.

Acknowledgments

We thank Zisu Mao and Jane M.C. Chan for technical assistance with our proteomics studies using two-dimensional gel electrophoresis and mass spectrometry respectively.

Author Contributions

Conceived and designed the experiments: SR BH. Performed the experiments: SR BH JW. Analyzed the data: SR BH JW ML. Contributed reagents/materials/analysis tools: SR BH. Wrote the paper: SR BH.

References

- Siomi H, Siomi MC (2010) Posttranscriptional regulation of microRNA biogenesis in animals. *Mol Cell* 38: 323–332.
- Calin GA, Croce CM (2006) MicroRNA signatures in human cancers. *Nat Rev Cancer* 6: 857–866.
- Dai Y, Huang YS, Tang M, Lv TY, Hu CX, et al. (2007) Microarray analysis of microRNA expression in peripheral blood cells of systemic lupus erythematosus patients. *Lupus* 16: 939–946.
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281–297.
- Doench JG, Sharp PA (2004) Specificity of microRNA target selection in translational repression. *Genes Dev* 18: 504–511.
- Kloosterman WP, Plasterk RH (2006) The diverse functions of microRNAs in animal development and disease. *Dev Cell* 11: 441–450.
- Wang X, Ye L, Hou W, Zhou Y, Wang YJ, et al. (2009) Cellular microRNA expression correlates with susceptibility of monocytes/macrophages to HIV-1 infection. *Blood* 113: 671–674.
- Pfeffer S, Sewer A, Lagos-Quintana M, Sheridan R, Sander C, et al. (2005) Identification of microRNAs of the herpesvirus family. *Nat Methods* 2: 269–276.
- Kincaid RP, Burke JM, Sullivan CS (2012) RNA virus microRNA that mimics a B-cell oncomiR. *Proc Natl Acad Sci U S A* 109: 3077–3082.
- Hariharan M, Scaria V, Pillai B, Brahmachari SK (2005) Targets for human encoded microRNAs in HIV genes. *Biochem Biophys Res Commun* 337: 1214–1218.
- Ahluwalia JK, Khan SZ, Soni K, Rawat P, Gupta A, et al. (2008) Human cellular microRNA hsa-miR-29a interferes with viral nef protein expression and HIV-1 replication. *Retrovirology* 5: 117.
- Yeung ML, Bannasser Y, Myers TG, Jiang G, Benkirane M, et al. (2005) Changes in microRNA expression profiles in HIV-1-transfected human cells. *Retrovirology* 2: 81.
- Houzet L, Jeang KT (2011) MicroRNAs and human retroviruses. *Biochim Biophys Acta* 1809(11–12): 686–693.
- Gupta A, Nagilla P, Le HS, Bunney C, Zych C, et al. (2011) Comparative expression profile of miRNA and mRNA in primary peripheral blood mononuclear cells infected with human immunodeficiency virus (HIV-1). *PLoS One* 6: e22730.
- Pilakka-Kanthikeel S, Saiyed ZM, Napuri J, Nair MP (2011) MicroRNA: implications in HIV, a brief overview. *J Neurovirol* 17: 416–423.
- Bannasser Y, Le SY, Yeung ML, Jeang KT (2004) HIV-1 encoded candidate micro-RNAs and their cellular targets. *Retrovirology* 1: 43.
- Lin J, Cullen BR (2007) Analysis of the interaction of primate retroviruses with the human RNA interference machinery. *J Virol* 81: 12218–12226.
- Omoto S, Fujii YR (2005) Regulation of human immunodeficiency virus 1 transcription by nef microRNA. *J Gen Virol* 86: 751–755.
- Kaul D, Ahlawat A, Gupta SD (2009) HIV-1 genome-encoded hiv1-mir-H1 impairs cellular responses to infection. *Mol Cell Biochem* 323: 143–148.
- Klase Z, Winograd R, Davis J, Carpio L, Hildreth R, et al. (2009) HIV-1 TAR miRNA protects against apoptosis by altering cellular gene expression. *Retrovirology* 6: 18.
- Ouellet DL, Plante I, Landry P, Barat C, Janelle ME, et al. (2008) Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. *Nucleic Acids Res* 36: 2353–2365.
- Yeung ML, Bannasser Y, Watashi K, Le SY, Houzet L, et al. (2009) Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid. *Nucleic Acids Res* 37: 6575–6586.
- Schopman NC, Willemsen M, Liu YP, Bradley T, van Kampen A, et al. (2012) Deep sequencing of virus-infected cells reveals HIV-encoded small RNAs. *Nucleic Acids Res* 40: 414–427.
- Hakim ST, Alsayari M, McLean DC, Saleem S, Addanki KC, et al. (2008) A large number of the human microRNAs target lentiviruses, retroviruses, and endogenous retroviruses. *Biochem Biophys Res Commun* 369: 357–362.
- Rasheed S, Yan JS, Lau A, Chan AS (2008) HIV replication enhances production of free fatty acids, low density lipoproteins and many key proteins involved in lipid metabolism: a proteomics study. *PLoS One* 3: e3003.
- Rasheed S, Yan JS, Hussain A, Lai B (2009) Proteomic characterization of HIV-modulated membrane receptors, kinases and signaling proteins involved in novel angiogenic pathways. *J Transl Med* 7: 75.
- Gong TW, Besirli CG, Lomax MI (2001) MACF1 gene structure: a hybrid of plectin and dystrophin. *Mamm Genome* 12: 852–861.
- Wang W, Guo J, Yu D, Vorster PJ, Chen W, et al. (2012) A Dichotomy in Cortical Actin and Chemotactic Actin Activity between Human Memory and Naive T Cells Contributes to Their Differential Susceptibility to HIV-1 Infection. *J Biol Chem* 287: 35455–35469.
- Boeras DI, Hraber PT, Hurlston M, Evans-Strickfaden T, Bhattacharya T, et al. (2011) Role of donor genital tract HIV-1 diversity in the transmission bottleneck. *Proc Natl Acad Sci U S A* 108: E1156–1163.
- Flavin RJ, Smyth PC, Laios A, O'Toole SA, Barrett C, et al. (2009) Potentially important microRNA cluster on chromosome 17p13.1 in primary peritoneal carcinoma. *Mod Pathol* 22: 197–205.
- Sun G, Rossi JJ (2011) MicroRNAs and their potential involvement in HIV infection. *Trends Pharmacol Sci* 32: 675–681.
- Li D, Zhao Y, Liu C, Chen X, Qi Y, et al. (2011) Analysis of MiR-195 and MiR-497 expression, regulation and role in breast cancer. *Clin Cancer Res* 17: 1722–1730.
- Zhu H, Yang Y, Wang Y, Li J, Schiller PW, et al. (2011) MicroRNA-195 promotes palmitate-induced apoptosis in cardiomyocytes by down-regulating Sirt1. *Cardiovasc Res* 92: 75–84.
- Zhu HC, Wang LM, Wang M, Song B, Tan S, et al. (2012) MicroRNA-195 downregulates Alzheimer's disease amyloid-beta production by targeting BACE1. *Brain Res Bull* 88(6): 596–601.
- Chen H, Untiveros GM, McKee LA, Perez J, Li J, et al. (2012) Micro-RNA-195 and -451 regulate the LKB1/AMPK signaling axis by targeting MO25. *PLoS One* 7: e41574.
- Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120: 15–20.
- Bannasser Y, Yeung ML, Jeang KT (2006) HIV-1 TAR RNA subverts RNA interference in transfected cells through sequestration of TAR RNA-binding protein, TRBP. *J Biol Chem* 281: 27674–27678.
- Pantophlet R, Burton DR (2006) GP120: target for neutralizing HIV-1 antibodies. *Annu Rev Immunol* 24: 739–769.
- Cheng-Mayer C, Quiroga M, Tung JW, Dina D, Levy JA (1990) Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation. *J Virol* 64: 4390–4398.
- Morikita T, Maeda Y, Fujii S, Matsushita S, Obaru K, et al. (1997) The V1/V2 region of human immunodeficiency virus type 1 modulates the sensitivity to neutralization by soluble CD4 and cellular tropism. *AIDS Res Hum Retroviruses* 13: 1291–1299.
- van Gils MJ, Bunnik EM, Boeser-Nunink BD, Burger JA, Terlouw-Klein M, et al. (2011) Longer V1V2 region with increased number of potential N-linked glycosylation sites in the HIV-1 envelope glycoprotein protects against HIV-specific neutralizing antibodies. *J Virol* 85: 6986–6995.
- Rasheed S, Norman GL, Gill PS, Meyer PR, Cheng L, et al. (1986) Virus-neutralizing activity, serologic heterogeneity, and retrovirus isolation from homosexual men in the Los Angeles area. *Virology* 150: 1–9.
- Meissner EG, Coffield VM, Su L (2005) Thymic pathogenicity of an HIV-1 envelope is associated with increased CXCR4 binding efficiency and V5-gp41-dependent activity, but not V1/V2-associated CD4 binding efficiency and viral entry. *Virology* 336: 184–197.
- Antonov AV, Dietmann S, Wong P, Lutter D, Mewes HW (2009) GeneSet2miRNA: finding the signature of cooperative miRNA activities in the gene lists. *Nucleic Acids Res* 37: W323–328.
- Karlin S, Altschul SF (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc Natl Acad Sci U S A* 87: 2264–2268.
- Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. *Comput Appl Biosci* 5: 151–153.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.