Published in final edited form as:

Biosystems. 2010 August ; 101(2): 88–96. doi:10.1016/j.biosystems.2010.05.001.

HIV-miR-H1 evolvability during HIV pathogenesis

Susanna L. Lamers^a, Gary B. Fogel^b, and Michael S. McGrath^{c,*}

^aBiolInfoExperts, Thibodaux, LA 70302, USA

^bNatural Selection, Inc., San Diego, CA 92121, USA

^cDepartment of Laboratory Medicine, Positive Health Program, University of California, San Francisco, CA 94110, USA

Abstract

The discovery of microRNAs (miRNAs) in viruses has generated considerable attention into their functional relevance in processes such as cell death, viral proliferation, and oncogenesis. Two early studies found no detectable miRNAs expressed within HIV; however, several studies have verified the existence and function of three HIV miRNAs, most notably HIV-miR-TAR, thus making the earlier results controversial. Although miRNAs are highly conserved within most species, HIV is known to have a high mutation rate, which could contribute to the opposing experimental findings and raises questions about whether all HIV miRNAs are robust enough to maintain their integrity, especially in viral regions prone to insertions and deletions. In addition, could the evolvability of HIV miRNAs contribute to the diversity in HIV disease pathogenesis? To address this question, we examined mutations in 1293 sequences in a suspect HIV miRNA, called miR-H1, derived from a large variety of tissues from seven patients. We found considerable diversity within the structures, including a patient-specific deletion and the potential for the development of new miRNAs as a result of deletions. We also note a potential disease association between a less stable miR-H1 and the development of AIDS-related lymphoma (ARL).

Keywords

MicroRNA; HIV; Evolution; Pathogenesis

1. Background

MicroRNAs (miRNAs) are regulators of gene expression and are involved in a variety of important cell functions including control of cell proliferation, apoptosis, and differentiation. miRNAs depend upon complementary hybridization to target sequences. Several functional miRNAs have been identified in human viruses, including Epstein-Barr virus, herpes viruses, and human cytomegalovirus (Pfeffer et al., 2004; Cai et al., 2006; Grey et al., 2007). The study of viral miRNAs is relatively new, and it appears that they are involved in both the modulation of the immune response as well as oncogenesis (Swaminathan, 2008). Three functional miRNAs have been identified within HIV-1, and three others have been proposed (Bennasser et al., 2004). Two earlier studies were unable to identify any miRNAs within HIV (Pfeffer et al., 2005; Lin and Cullen, 2007); however, these studies are controversial considering subsequent papers that have successfully identified HIV miRNAs, in particular, the well-characterized HIV miRNA, miR-TAR (Klase et al., 2007, 2009; Ouellet et al., 2008). Experimentally identifying HIV miRNAs could prove complicated if the miRNAs

were expressed at different stages of infection, within different infected cell-types, or expressed at very low levels. Additionally, these viruses may express miRNAs that are difficult to identify using the small-RNA cloning protocols described by Pfeffer et al. (2005) and Lin and Cullen (2007). Furthermore, HIV could potentially contain miRNAs in regions where the virus is both robust and evolvable.

The TAR structure is a 59-nucleotide stem-bulge-loop located in the 5' end of HIV transcripts both in the nucleus and the cytoplasm. TAR serves as a target for viral transcription by the TAT protein, is essential for efficient viral transcription and functions in the cytoplasm to inhibit translation. A second HIV miRNA, miR-N367, is a 70-nucleotide structure located in the nef/LTR overlap of the HIV genome (Omoto et al., 2004; Omoto and Fujii, 2005). Omoto and Fujii (2005) demonstrated that miR-N367 could reduce HIV-1 promoter activity and deduced that the intact miRNA down-regulates HIV transcription. A third HIV-1 miRNA, miR-H1, is an 81-nucleotide stem-loop structure located immediately downstream of the two NF- κ B sites in the LTR and has the ability to cleave the 12th exon of the apoptosis antagonizing transcription factor (AATF) gene resulting in degradation of that gene product (Kaul et al., 2006, 2009). Down-regulation of AATF expression is accompanied by a reduction in cell viability and down-regulation of dicer, which plays a crucial role in providing immunity at the nucleic acid level (Passananti and Fanciulli, 2007). The maintenance of the AATF gene may also be important in initiating cellular proliferation, repression of cellular apoptosis, and maintenance of the viral reservoir (Xie and Guo, 2004). Others have speculated that AATF may be a component of the anticancer barrier that protects cells from DNA damage or oncogenic stress (Bartkova et al., 2005; Gorgoulis et al., 2005; Passananti and Fanciulli, 2007).

Because the function of HIV-1 miR-H1 has only been examined in the context of blood mononuclear cells, we sought to determine the integrity of the miR-H1 structure in HIV-1 sequences derived from tissues where macrophages were a prominent HIV-infected cell type. Such sequences included those derived from brain tissues of patients who developed HIV-associated dementia (HAD) and tumor tissues from patients who developed AIDS-related lymphoma (ARL) (Zenger et al., 2002; Salemi et al., 2009; Lamers et al., 2009). Macrophages are considered long-lived reservoirs of HIV during infection, contributing to both HAD and ARL development. These diseases are still prevalent despite current antiviral therapies. Interestingly, HAD was dominant at the beginning of the HIV epidemic, whereas ARL did not evolve as an AIDS-associated disease until the mid-1980s. In addition, it is also known that different subtypes of HIV are associated with varied prevalence of AIDS-associated diseases (Liner et al., 2007; Sacktor et al., 2009). To date, no known genetic determinant for variation within HIV-related disease processes has been discovered.

Human cellular miRNA expression can change dramatically between tissues. To our knowledge ours is the first study to examine HIV miRNA sequence variation within different categories of HIV-associated diseases and across multiple tissues. Kaul et al. (2009) showed that the variation in the mature HIV-1 miR-H1 is found in a variety of HIV subtypes; however, they did not examine if this variation could alter the pre-miRNA structures (Kaul et al., 2009). In this study we examined the conservation of both the mature HIV-1 miR-H1 sequence and its associated pre-miRNA structure in 1293 sequences derived from 55 post-mortem tissues from 7 individuals who died from a variety of complications following HIV infection. Preliminary disease associations to the structures of HIV-1 the miR-H1 were assessed.

2. Methods

2.1. Tissues

Extensive details concerning patient tissue samples utilized in the current study have been published previously (Salemi and Goodenow, 2008). Briefly, patients AM and IV developed metastatic ARL, patients GA and CX developed progressive HAD, patient BW developed lymphoma, including meningeal lymphoma and HAD, patient DY had suffered from recurring *Mycobacterium avium* complex infection (MAC) for years and developed an aggressive form of HAD in the months prior to death and patient AZ died due to severe cardiovascular disease. Multi-site frozen autopsy specimens of the tissues necessary for this study were obtained through the AIDS and Cancer Specimen Resource (ACSR) (<http://acsr.ucsf.edu>) after appropriate consent and the application of a de-identification procedure. Patient two-letter designations used throughout this study were generated randomly as shorthand used by technicians who performed the studies. The ACSR is recognized by the Office of Biorepositories and Biospecimen Research at the National Institutes of Health as being HIPPA compliant and in accordance with the ethical standards of the Declaration of Helsinki.

2.2. Amplification, Cloning and Sequencing

Amplification, cloning, and sequencing of the gp120-LTR virus DNA were obtained using primers and conditions previously described (Lamers et al., 2009). Approximately 20–30 sequences were derived from each tissue used in the study. Tissues from each patient are described in Tables 4–10.

2.3. Data Analysis

For distance analysis, we used the 560-nucleotide region, relative to HXB2 coordinates 1–560, that incorporates the regulatory domain of the LTR. The analysis was performed in MEGA (Tamura et al., 2007) using the Tamura 3-Parameter model with gapped regions removed.

The HIV-1 pre-miR-H1 sequence was identified in each patient's aligned LTR data set using MEGA (Tamura et al., 2007). The alignment was reduced to contain non-identical sequences using DAMBE (Xia and Xie, 2001). For each sequence group a consensus structure was derived using the published structure as a guide and covariation analysis (Fig. 1). An optimized sequence alignment of the patient's data would, in certain regions, introduce gaps in different places than shown in the pre-miRNA structures; however, gaps were introduced to the structure only to maximize stability. The program mfold (Zuker, 2003) was used to determine the folding energies of the H1 miRNAs at 37 °C. Stem-loop structures for 10 common molecular clone sequences (JRCSF, ADA, YU2, NL43, MN, LAI, 89.6, SF2, DH12, WR27) with known co-receptor phenotypes (CCR5, CXCR4 or dual-tropic) were constructed and are shown in Supplemental Figure 1 (S1). BLAST (Benson et al., 1997) was used to determine the relative prevalence of each mature HIV-1 miR-H1 in the National Center for Biotechnology Information GenBank database.

Patient IV lacked the mature miRNA in all cases, and as a result the sequence information directly 5' of the deletion was used to construct a potentially new HIV miRNA structure (boxed region in Fig. 4). Patient CX possessed two sequences derived from periventricular brain tissue that were missing either the 5' or 3' sequence of the stem-loop structure (boxed regions in Fig. 5). Sequence data immediately 5' or 3' of this deletion was used to attempt to construct a new miRNA structure. The folding energies were calculated for each putative new miRNA structure are shown in Table 3 along with %GC content, the number of paired and unpaired nucleic acids in the stem-loop structure, and the size of the hairpin loop.

A nucleotide substitution matrix was developed for each patient by counting each nucleotide that varied from Group 1 (Fig. 7). The probabilities were calculated as substitution type/total number of substitutions over the pre-miRNA structure.

3. Results

Distance analysis showed that the population of sequences between patients could vary from 5.5 to 11.9% with a median of 8.7% (Fig. 1). Within patient sequence divergence was much lower, ranging from 1.2 to 3.9% with a median of 2.2% (Fig. 2). A *t*-test showed that the sequence divergence between patients was significantly higher than the sequence divergence within patient sequence populations ($p < 0.0001$).

Each patient contained a dominant pre-miRNA sequence for pre-miR-H1 that is shown in Table 1. In Table 2, the dominant mature HIV-1 miR-H1 sequences are provided with variable positions indicated. In Tables 4–10, we provide a summary of all the pre-miR-H1s for each patient as follows: (1) Group 1 is the dominant group, (2) Groups 2–11 contained three or more sequences, and (3) the final group for each patient contains all sequences with two or less representative stem-loop structures. The number of sequences in each group, along with the number and type of tissues the sequences were derived from are also shown. In the lymphoma patients (AM and IV), tissues are further classified as tumor, non-tumor, or lymph nodes. In the patients where sequences from brain tissues were available (BW, AZ, DY, CX, GA), the tissues are also identified as brain or non-brain.

Patients with brain infection or clinical HAD (GA, CX, DY, BW) shared similar miR-H1 structures (Fig. 3). Additionally, eight of the 10 miR-H1 structures built from molecular clone sequences contained structures similar to GA, CX, DY, and BW in that each had a similar length, no apparent deletion or insertion and a similar loop (Supplemental Figure 2). The structure for patient AZ, a patient with cardiovascular disease (CVD), was longer than all other miRNAs examined by 12 nucleotides, most of which were located in the terminal loop and away from the mature miRNA sequence. Two patients with ARL contained much different miR-H1 structures. Patient AM exhibited a very large bulge in the center of the stem that contained a significant proportion of the mature miRNA sequence. Patient IV no longer has the mature miR-H1 in any sequences. Fig. 4 illustrates that despite this deletion, patient IV's pre-miR structure can be reconstituted if additional sequence data upstream of the deletion is used to complete the structure. A similar approach was taken for two sequences derived from patient CX's periventricular space that also exhibited deletions within the HIV-1 miR-H1 (Fig. 5). All three structures are consistent with a viable pre-miR structure, especially the two structures derived from the CX deletions.

Fig. 6 shows substitution patterns within the stem-loop structure for each patient and each of their non-identical sequences. Each nucleic acid substitution is shown along with its associated group number. Substitutions that could alter the Group 1 pre-miR-H1 structure are highlighted in red. A substitution matrix developed for each patient clearly indicates a bias towards $G \rightarrow A$ transitions within the overall data set ($n = 161$), followed by $U \rightarrow C$ ($n = 79$) an almost equal amount of $A \rightarrow G$ ($n = 42$) and $C \rightarrow U$ ($n = 39$) transitions. As would be expected in the paired regions of a stable RNA structure, transversions were rare. Substitution matrices for all patients are provided as Supplemental Figure (S2). A substitution matrix compiled from all patients is shown in Fig. 7.

Alignment of the stem-loop sequences showed three nucleic acid insertions in our data that had not been reported previously in the published sequence. These positions are highlighted in Tables 1 and 2. The first insertion, either an adenine or guanine, occurred within the mature miRNA sequence. BLAST revealed that this insertion was found in most HIV-1 LTR sequences (>6000 sequences) within GenBank. These results suggest that the published

miR-H1 mature sequence is uncommon and not likely to be the majority structure for this miRNA family. The second insertion, an AG, occurred at the 3' end of the loop structure. The insertion was also highly represented in the GenBank sequences; however, it is unlikely that the insertion of AG nucleotides in the loop would significantly alter the stability of the pre-miRNA structure.

The stability of the miRNA stem-loop secondary structure can be estimated by (1) calculating the energy needed for folding (the lower the energy, the more likely a miRNA structure will form), (2) calculating the GC content of the structure (the higher the GC content, the more likely the region is to be stemmed), (3) observing the number of paired and unpaired nucleotides in the stem and, (4) observing the size of the hairpin loop (less than 5 or more than 10 nucleotides form less stable structures) (Griffiths-Jones et al., 2003; Lim et al., 2003; Pfeffer et al., 2004). A comparison of these parameters for each of the dominant sequences is presented in Table 3. Three of our sequences had a lower folding energy than the previously published miR-H1 structure (sequences from patients AZ, CX-1B, and CX-2B). Patient AZ's decreased folding energy is likely due to the increased size of the structure and the increased loop length. It is interesting that the putative H1 miRNA structures constructed using data immediately adjacent to the deletions for patient CX appeared to be quite stable (Fig. 5 and Table 3). One of patient CX's structures contains the mature miRNA sequence (CX-2B) and the other does not (CX-1B). The new miRNA for patient IV has a low folding energy, but contains a large number of unpaired nucleotides, therefore it is unclear how stable this structure would be in nature. The miRNA for patient AM is the least stable miRNA structure found in our data set.

A dominant pre-miR-H1 sequence usually appeared in most, if not all, of the sampled patient's tissues (Tables 4–10), suggesting that the HIV miRNA structure for each patient generally did not mutate due to tissue-selective immune pressures and that there was selection for maintenance of specific structures within each patient. This hypothesis is supported by the fact that variants differing from the dominant pre-miR-H1 structure in patients were rare. Exceptions included non-tumor stomach sequences from patient IV, which contained many adenine substitutions (highlighted in green in Fig. 4). A phylodynamic study (Salemi et al., 2009) also determined that this patient's stomach sequences were highly divergent from tumor and lymph node sequences. Nine liver-derived sequences in patient DY (Group 3) also contained a structure with multiple adenine substitutions (highlighted in green in Fig. 6). These nine sequences would result in major structural changes in the HIV-1 pre-miR-H1 and generate the highest folding energy of all sequences examined (−25.50 kcal/mol). Temporal lobe sequences from patient GA and frontal lobe white matter from patient BW were not found in the dominant group; however the changes from the dominant sequence were slight and did not alter the structure.

Fig. 6 shows all of the variation in each patient over the new HIV-1 pre-miR-H1 consensus structure. Substitutions in the center of the stem structure opposite of the mature miRNA sequence were rare, found only randomly in individual sequences (motif = CGUCGACGAAUAUACGUCGUAGACUCCCGA). Substitutions within the mature HIV-1 miR-H1 sequence were found more frequently in the patients as follows: GA = 5 groups, BW = 3 groups, CX = 5 groups, DY = 4 groups, AZ = 2 groups, AM = 8 groups, IV = not applicable due to deletion. The increased number of substitutions within the AM mature miRNA sequence coincides with the decreased stability within this patient's miRNA. Overall, most substitutions would not change the structure of the miRNA. However, as Bennasser et al. (2004) noted, the identification of a more stable sequence opposite of the known miRNA could indicate that an additional mature miRNA sequence might exist in the structure (Bennasser et al., 2004). As with the published pre-miR-H1 sequence, the

conserved motif on the opposite of the stem loop had numerous human mRNA targets when scanned against the mRNA database at NCBI (data not shown).

4. Discussion

Human viruses may contain miRNAs that modulate the expression of host genes (Pfeffer, 2007). This is accomplished by their complementarity and subsequent repression of specific human mRNAs. Complementarity can be imperfect and yet still functional; therefore, the identification of miRNA targets can be particularly difficult to predict (Provost et al., 2006). Within HIV, the high rate of variability within and between patients increases the sequence reservoir of miRNAs and adds another level of complexity in the identification of viable miRNA structures. Although only three HIV miRNAs have been experimentally examined, bioinformatic approaches have identified at least three additional miRNA candidates (Bennasser et al., 2004). In this study, we have also uncovered additional potential HIV miRNAs from the miR-H1 family.

Through the examination of HIV miR-H1 in a large cohort of tissues sampled from seven individuals who died from a variety of macrophage-associated and AIDS-related disease processes, several lessons can be learned. The nature of the HIV-1 LTR regulatory domain between patients may provide HIV with great phenotypic plasticity, and hence increased evolvability. The variability of HIV pre-miR-H1 sequences in tissues from HIV-infected individuals was larger than anticipated and differed from the published structure in every sequence examined. The prior pre-miR-H1 structure proposed by Kaul et al. (2009) suffered from alignment errors that masked the insertion of a nucleic acid in nearly all of the published mature miRNA sequences besides the published H1 structure (Kaul et al., 2009). This observation highlights the importance of deriving miRNA structures from multiple sequences and calls into question any miRNA structures that are hypothesized from one or a handful of sequences and then placed into databases such as the miR Registry (Griffiths-Jones, 2004). It is unfortunately likely that other viral miRNAs in the miR Registry demonstrate significant variation, although this has yet to be confirmed.

The new consensus pre-miR-H1 structure was reasonable for all but two patients and in two sequences from an additional patient. This very interesting phenomenon suggests that even within a single patient, HIV-associated miRNAs may vary considerably and may even be deleted. To our knowledge this is the first time that viral miR deletion has been identified in a patient-specific manner. The instability of the AM structure, the loss of the H1 miRNA in all of patient IV sequences, and the ability of these regions to form miRNA-like structures despite large deletions in known miRNAs raises several interesting questions: (1) would the deletion of HIV-miR-H1 have the opposite effect on the AATF pathway, leading to cellular proliferation of HIV-infected cells, rather than cell death? (2) Could the new miRNAs generated due to deletion events be viable and have new targets? (3) Do deletions or insertions in other regions of the HIV U3' LTR generate presently unknown HIV miR-NAs? Considering the large amount of variability within the HIV genome, it seems likely that significant variation occurs within HIV miRNAs and may be part of a co-evolutionary dynamic between the host and virus miRNA during the infection process. Such variation in the HIV miRNA may affect disease progression, and the ability of the host environment to alleviate such progression.

In both cases where significant alteration of the HIV miR-H1 existed the patients developed ARL. According to several studies, antigenic stimulation of B-cells by activated macrophages plays a central role in the pathogenesis of ARL (Herndier et al., 1994; McGrath, 1997; Ng and McGrath, 1998). If the AATF pathway is not being affected negatively by HIV-1 miR-H1, as proposed by Kaul et al. (2009), this may provide an

additional means for activated macrophages to avoid apoptosis and produce factors that could stimulate lymphoma growth. Other mechanisms for macrophages to avoid apoptosis have also been suggested (Swingler et al., 2007). The knockdown of the AATF gene is associated with suppression of important cell regulators Par-4 and c-myc. Conversely, Par-4 activation leads to the production of various cytokines and chemokines that can modulate immune and inflammatory responses (Rattenholl and Steinhoff, 2008). C-myc is involved in cell cycle and growth. In patients IV and AM, the variability of the HIV-1 miR-H1 did not result in an ineffective virus population; however, it could have perturbed the AATF or other pathways in ways yet unknown. The observation that data from two ARL patients had non-existent or suspect HIV-1 miR-H1 sequences is intriguing; however, more LTR sequences generated from lymphoma tissues and other experimental studies would be necessary to confirm if HIV-1 miR-H1 has any relationship to the specific disease.

All patients with HAD shared an almost identical miR-H1 structure. This structure was similarly represented in 8 of the 10 HIV molecular clone sequences examined and appeared the most related to the original published miR-H1 sequence. Although ARL and HAD are both macrophage-related disease processes, they differ in that ARL develops due to crosstalk between macrophages and B-cells, whereas HAD is an inflammatory process and related to the accumulation of HIV-infected macrophages in the brain. Xie et al. observed that AATF is expressed in cortical neurons and protects them from oxidative stress (Xie and Guo, 2004). Although HIV does not infect neurons directly, it is possible that miRNAs that specifically target the AATF pathway could be involved in HAD development. The finding of potentially different miRNAs in ARL vs. HAD patients is noteworthy in that it could potentially lead to new hypothesis concerning the development of varied macrophage-associated diseases.

Because ARL is a disease that has increased in prevalence during the HIV epidemic, the identification of evolved or deleted miRNAs within the viral genome of ARL patients is significant discovery. In addition, examination of the sequence data revealed that the published structure of miR-H1 is uncommon, and our data supports a new consensus structure that is generated over a large sequence set.

The discovery of HIV miRNAs with large deletions and the determination that mutations in HIV miRNAs over many tissues and patients generate a diverse set of pre and mature miRNAs, even for one miRNA, has important ramifications for experimental verification and miRNA evolution. In addition, the large amount of variation found in HIV miRNAs suggests that HIV may be a good model for the examination of miRNA evolution and also tests of evolvability.

The extent of HIV miR variability has not been considered properly in the current literature. In fact, conservation of a miRNA is the standard logical approach to determine if a pre-miRNA structure is supportive of a viable mature miRNA sequence. However, variation due to defects in HIV reverse transcriptase is a normal attribute of HIV evolution and, moreover, has resulted in the successful establishment of various HIV subtypes and phenotypes (Kirchhoff et al., 1997; Skinner et al., 1998; Aquino-De Jesus et al., 2000; Briggs et al., 2000; Hoffman et al., 2002; Dyer et al., 2008). The standard convention in the field of miRNA research has been to take a very conservative view and name miRNA families based on sequence similarity, primarily in the seed region of the first 2–7 nucleotides of the mature microRNA. This convention works well when the seed region is conserved. Despite the variability associated with HIV-miR-H1, the seed region is often still preserved, and provides the opportunity to preserve some of the phylogenetic history of its evolution. Thus, HIV may be using its extremely high mutation rate to explore a much wider distribution of microRNA sequences than most eukaryotes do. It stands to reason that this level of

evolvability is another way for the virus to combat the defenses of the human cell, and increase the probability of its success during infection. Such indications await future experimentation. Therefore, current approaches to identify HIV miRNA functions should also consider the evolvability of miRNA as a reasonable contributor to the large spectrum of immune responses seen in HIV-infected patients. In this study, we examined genomic data and determined that variation can potentially change the structure of an HIV-associated miRNA or generate new miRNAs. Total RNA expression in these tissues awaits further investigation. The possibility exists that understanding HIV miRNA evolvability may result in new hypotheses concerning HIV progression of disease or to long-term non-progression. Additionally, in light of the fact that evolution of miRNAs across species is a rapidly expanding field, the variability found within the HIV miRNAs found in this study highlights that viruses, specifically HIV, may be a worthwhile species in which to study miRNA evolution.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The project was funded by the National Institutes of Health grants U01 CA096230-06 and U01 CA066529-14. The authors would like to thank the individuals who have assisted in the generation and formatting of the data set: Li Zhao, Derek Galligan, Alanna Morris, Marco Salemi, Tulio de Oliveira, Sara Granier and the general support of the individuals at the West Coast ACSR.

References

- Aquino-De Jesus MJ, Anders C, et al. Genetically and epidemiologically related “non-syncytium-inducing” isolates of HIV-1 display heterogeneous growth patterns in macrophages. *Journal of Medical Virology*. 2000; 61 (2):171–180. [PubMed: 10797371]
- Bartkova J, Horejsi Z, et al. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature*. 2005; 434 (7035):864–870. [PubMed: 15829956]
- Bennasser Y, Le S, et al. HIV-1 encoded candidate micro-RNAs and their cellular targets. *Retrovirology*. 2004; 1:43. [PubMed: 15601472]
- Benson DA, Boguski MS, et al. GenBank. *Nucleic Acids Research*. 1997; 1 (25):1–6. [PubMed: 9016491]
- Briggs DR, Tuttle DL, et al. Envelope V3 amino acid sequence predicts HIV-1 phenotype (co-receptor usage and tropism for macrophages). *AIDS*. 2000; 14 (18):2937–2939. [PubMed: 11153675]
- Cai X, Schafer A, et al. Epstein-Barr virus microRNAs are evolutionarily conserved and differentially expressed. *PloS Pathogens*. 2006; 2 (3):e23. [PubMed: 16557291]
- Dyer WB, Zaunders JJ, et al. Mechanisms of HIV non-progression; robust and sustained CD4⁺ T-cell proliferative responses to p24 antigen correlate with control of viraemia and lack of disease progression after long-term transfusion-acquired HIV-1 infection. *Retrovirology*. 2008; 11 (5):112. [PubMed: 19077215]
- Gorgoulis VG, Vassiliou LVF, et al. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature*. 2005; 434:907–913. [PubMed: 15829965]
- Grey F, Meyers H, et al. A human cytomegalovirus-encoded microRNA regulates expression of multiple viral genes involved in replication. *PloS Pathogens*. 2007; 3 (11):e163. [PubMed: 17983268]
- Griffiths-Jones S. The microRNA registry. *Nucleic Acids Research*. 2004; 32:D109–D111. [PubMed: 14681370]
- Griffiths-Jones S, Bateman A, et al. Rfam: an RNA family database. *Nucleic Acids Research*. 2003; 31 (1):239–241.

- Herndier BG, Kaplan L, et al. Pathogenesis of AIDS lymphomas. *AIDS*. 1994; 8 (8):1025–1049. [PubMed: 7986399]
- Hoffman NG, Seiller-Moiseiwitsch F, et al. Variability in the human immunodeficiency virus type 1 gp120 Env protein linked to phenotype-associated changes in the V3 loop. *Journal of Virology*. 2002; 76 (8):3852–3864. [PubMed: 11907225]
- Kaul D, Ahlawat A, et al. HIV-1 genome-encoded hiv-1mir-H1 impairs cellular responses to infection. *Molecular and Cellular Biochemistry*. 2009; 323:123–128.
- Kaul D, Khanna A, et al. Evidence and nature of a novel miRNA encoded by HIV-1. *Proceedings of the Indian National Science Academy*. 2006; 72:91–95.
- Kirchhoff F, Greenough TC, et al. Activity of human immunodeficiency virus type 1 promoter/TAR regions and tat1 genes derived from individuals with different rates of disease progression. *Virology*. 1997; 232 (2):319–331. [PubMed: 9191845]
- Klase Z, Kale P, et al. HIV-1 TAR element is processed by Dicer to yield a viral micro-RNA involved in chromatin remodeling of the viral LTR. *BMC Molecular Biology*. 2007; 8:63. [PubMed: 17663774]
- Klase Z, Winograd R, et al. HIV-1 TAR miRNA protects against apoptosis by altering cellular gene expression. *Retrovirology*. 2009; 6:18. [PubMed: 19220914]
- Lamers SL, Salemi M, et al. Extensive HIV-1 intra-host recombination is common in tissues with abnormal histopathology. *PLoS ONE*. 2009; 4 (3):e5065. [PubMed: 19333384]
- Lim LP, Glasner ME, et al. Vertebrate microRNA genes. *Science*. 2003; 299 (5612):1540. [PubMed: 12624257]
- Lin J, Cullen BR. Analysis of the interaction of primate retroviruses with the human RNA interface machinery. *Journal of Virology*. 2007; 81:12218–12226. [PubMed: 17855543]
- Liner KJ, Hall CD, et al. Impact of human immunodeficiency virus (HIV) subtypes on HIV-associated neurological disease. *Journal of Neurovirology*. 2007; 13 (4):291–304. [PubMed: 17849313]
- McGrath MS. T-cells and macrophages in HIV disease. *Clinical Reviews in Allergy and Immunology*. 1997; 14 (4):359–366. [PubMed: 9040966]
- Ng VL, McGrath MS. The immunology of AIDS-associated lymphomas. *Immunological Reviews*. 1998; 162:293–298. [PubMed: 9602372]
- Omoto S, Fujii YR. Regulation of human immunodeficiency virus 1 transcription by nef microRNA. *Journal of General Virology*. 2005; 86:751–755. [PubMed: 15722536]
- Omoto S, Ito M, et al. HIV-1 nef suppression by virally encoded microRNA. *Retrovirology*. 2004; 1:44. [PubMed: 15601474]
- Ouellet DL, Plante I, et al. Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. *Nucleic Acids Research*. 2008; 36:2353–2365. [PubMed: 18299284]
- Passananti C, Fanciulli M. The anti-apoptotic factor Che-I/AATF links transcriptional regulation, cell cycle control, and DNA damage response. *Cell Division*. 2007; 16 (2):21. [PubMed: 17634135]
- Pfeffer S. Micro RNA and viral infections in mammals. *Journal de la Societe de biologie*. 2007; 201 (4):377–384. [PubMed: 18533098]
- Pfeffer S, Sander C, et al. Identification of microRNAs of the herpesvirus family. *Nature Methods*. 2005; 2:269–276. [PubMed: 15782219]
- Pfeffer S, Zavolan M, et al. Identification of virus-encoded microRNAs. *Science*. 2004; 304 (5671):734–736. [PubMed: 15118162]
- Provost P, Barat C, et al. HIV-1 and the microRNA-guided silencing pathway. *Virus Research*. 2006; 121:107–115. [PubMed: 16889864]
- Rattenholl A, Steinhoff M. Proteinase-activated receptor-2 in the skin: receptor expression, activation and function during health and disease. *Drug News Perspective*. 2008; 21 (7):369.
- Sacktor N, Nakasujja N, et al. HIV subtype D is associated with dementia compared to subtype A in immunosuppressed individuals at risk for cognitive impairment in Kampala, Uganda. *Clinical Infectious Diseases*. 2009; 49 (5):780–786. [PubMed: 19622045]
- Salemi M, Goodenow MM. An exploratory algorithm to identify intra-host recombinant viral sequences. *Molecular Phylogenetics and Evolution*. 2008; 49 (2):618–628. [PubMed: 18801446]

- Salemi M, Lamers SL, et al. Distinct patterns of HIV-1 evolution within metastatic tissues in patients with non-Hodgkins lymphoma. *PLoS ONE*. 2009; 4 (12):e8153. [PubMed: 19997510]
- Skinner LM, Lamers SL, et al. Analysis of a large collection of natural HIV-1 integrase sequences, including those from long-term nonprogressors. *Journal of Acquired Immune Deficiency Syndromes*. 1998; 19 (2):99–110.
- Swaminathan S. Noncoding RNAs produced by oncogenic human her-pesviruses. *Journal of Cellular Physiology*. 2008; 216 (2):321–326. [PubMed: 18484093]
- Swingler S, Mann AM, et al. Apoptotic killing of HIV-1-infected macrophages is subverted by the viral envelope glycoprotein. *PLoS Pathogens*. 2007; 3 (9):1281–1290. [PubMed: 17907802]
- Tamura K, Dudley J, et al. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*. 2007; 24:1596–1599. [PubMed: 17488738]
- Xia X, Xie Z. DAMBE: software package for data analysis in molecular biology and evolution. *The Journal of Heredity*. 2001; 92 (4):371–373. [PubMed: 11535656]
- Xie J, Guo Q. AATF protects neural cells against oxidative damage induced by amyloid beta-peptide. *Neurobiology of Disease*. 2004; 16 (1):150–157. [PubMed: 15207272]
- Zenger E, Abbey NW, et al. Injection of human primary effusion lymphoma cells or associated macrophages into severe combined immun-odeficient mice causes murine lymphomas. *Cancer Research*. 2002; 62 (19):5536–5542. [PubMed: 12359765]
- Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*. 2003; 31 (13):3406–3415. [PubMed: 12824337]

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.biosystems.2010.05.001.

Fig. 1. Distance between patient sequence populations. Each patient was given a color code as shown in the figure legend. The distance between the patient on the x -axis and other patients studied is shown as a colored bar. Distances are provided as a percent with standard errors displayed on each bar. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 2. Distance within patient sequence populations. Distances within each patient's sequences are provided as a percent with standard errors displayed on each bar. The number of sequences used for each patient is shown below the patient designation.

Fig. 3. Dominant stem-loop structures of the HIV-1 miR-H1 sequence for all patients. The published stem-loop structure is shown in the upper left of the figure. The mature miRNA sequence, which would be excised after dicer processing, is highlighted in blue. For each patient in the study, nucleotide changes from published structure are highlighted in red. Gaps introduced to maximize the structure's stability are shown with dashes. Hash marks are used to indicate bulges where a different number of nucleic acids occurred on either side of the stem. Base pairing is indicated with solid lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

IVb (5' deletion)

Fig. 4.

Replacement of the deletion in patient IV's stem-loop structure with data 3' to the deletion. Notation used for structures is identical to that in Fig. 1. The nucleotides 3' to the deletion, which became part of a potentially new miRNA, are boxed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

CX-1B (5' deletion)

CX-2B (3' deletion)

Fig. 5. Sequences in patient CX containing deletions. Notation used in structures is identical to that in Fig. 1. The nucleotides 3' or 5' to the deletion, which became part of a potentially new miRNA, are boxed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 6.

Nucleic acid substitution patterns for each patient over all sequence groups. Notation used for structures is identical to that in Fig. 1. Sequence groups were defined in Tables 4–10. Arrows point to substitutions. Substitutions are indicated using the notation “nucleic acid substitution:group number.” In some cases a substitution occurred over multiple groups and the group numbers are separated with a comma. An “X” indicates the loss and a “^” indicates the addition of a nucleic acid. Substitutions that would potentially alter the structure are shown in red. Patient IV contained significant variation in stomach tissue (Group 6) from other sequences in the data set. These substitutions are highlighted in green. Patient DY also contained significant variation in a subset of liver sequences (Group 3) and

these are also highlighted in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Nucleotide change from group 1

Nucleotide in group 1		G	A	U	C
	G	-	.45	.02	.01
	A	.12	-	0	.01
	U	.05	.01	-	.22
	C	.002	.01	.11	-

Fig. 7. Nucleotide substitution probability matrix. Substitutions and probabilities were calculated over all pre-miR-H1 genomes as described in Section 2. The probability of transitions and transversions are provided in the matrix.

Table 1

HIV-1 mir-H1 stem-loop sequences.

miRNA	Pre-miR-H1
H1	UCCAGGGAGGCGUGCCUGGGCGGGACUGGGGAGUGGCGCCUCAGAUCCUGCAUAUAAGCAGCGCUUUUUCGGUGCGUUA
GA	UCCAGGGAGGCGUG GC AUGGGCGGGACUGGGGAGUGGCG AG CCUCAGAUCCUGCAUAUAAGCAGCGCUUUUUCGGUGCGUUA
CX	UCCAGGGAGGCGUG GC CCUGGGCGGGACUGGGGAGUGGCG AG CCUCAGAUCCUGCAUAUAAGCAGCGCUUUUUCGGUGCGUUA
DY	UCCAGGGAGGCGUG GC CCUGGGCGGGACUGGGGAGUGGCG AG CCUCAGAUCCUGCAUAUAAGCAGCGCUUUUUCGGUGCGUUA
BW	UCCAGGGAGGCGUG GC CCUGGGCGGGACUGGGGAGUGGCG AG CCUCAGAUCCUGCAUAUAAGCAGCGCUUUUUCGGUGCGUUA
AM	UCCAGGGAGGCGUG GC CCUGGGCGGGACUGGGGAGUGGCG AG CCUCAGAUCCUGCAUAUAAGCAGCGCUUUUUCGGUGCGUUA
AZ	UCCAGGGAGGCGUG GC CCUGGGCGGGACUGGGGAGUGGCG AG CCUCAGAUCCUGCAUAUAAGCAGCGCUUUUUCGGUGCGUUA
IV	UCC-----GGUGGGCGGGACGGGGGAGUGGCG AG CCUCAGAUCCUGCAUAUAAGCAGCGCUUUUUCGGUGCGUUA
IVb	GGACUUUCCGCGGGACUUU CCGUGGGCGGGACGGGGGAGUGGCG AG CCUCAGAUCCUGCAUAUAAGCAGCGCUUUUUCGGUGCGUUA
CX-1B	UGGCCCCGAGAGCUGCAUCCGGCCU UGGGCGGGACUGGGGAGUGGCG AG CCUCAGAUCCUGCAUAUAAGCAGCGCUUUUUCGGUGCGUUA
CX-2B	GGACUUU UCCAGGGAGGCGUG GC CCUGGGCGGGACUGGGGAGUGGCG AG CCUCAGAUCCUGCAU UGACCU GGGAGCUCUCUGGCUAGCUAG

Bold letters indicate insertions found in all sequences that are not found in the published miR-H1 sequence. Dashes in the sequences for patient IV were introduced to indicate deletion in all of the patient's data. Red highlighted data indicates sequence data 5' or 3' of a deletion in all of IV sequences and two CX sequences.

Table 2

Aligned mature miR-H1 sequences.

miRNA	Maturemir-H1 RNA
	*
H1	CCAGGG - AGGCGUG - CCUGGGC
GA	CCAGGG - AGGCGUG GCA UGGGC
CX	CCAGGG - AGGCGUG GCC UGGGC
DY	CCAGGG - AGGCGUG GCC UGGGC
BW	CCA UGG GAGGCG CGG CCUGGGC
AM	CCAGGG - AGGCGUG GCCGGGGU
AZ	CCAGGG G AGGCGUG A CCUGGGC

Red highlighted nucleic acids indicate variation from H1. The asterisk indicates an insertion that is found in all sequences analyzed, including >8000 LTR sequences available in the NCBI sequence database.

Table 3

pre-miR-H1 secondary structure information.

miRNA	Folding energy (kcal/mol)	GC content (%)	Paired (nt)	Unpaired (nt)	Terminal loop length (nt)
H1	-33.40	81	29	8-10	5
GA	-31.70	78	24	9-11	5
CX	-30.20	80	25	8-10	5
DY	-32.80	80	26	8-10	5
BW	-28.80	80	27	7-9	5
AZ	-34.20	78	27	9-11	9
AM	-28.80	80	22	10-13	5
IV	NA	NA	NA	NA	5
IVb	-32.60	81	20-22	11-15	5
CX-1B	-35.10	80	27	10-12	5
CX-2B	-33.60	82	28	13	5

nt = nucleotides, NA = not applicable due to nucleotide deletions in the miRNA.

Table 4

Sequence groups for patient AM.

Group #	# of sequences	# of tissues	Tissues
1	176	10	Lymph nodes (24): AX-r, Ax-l Tumor (83): LI-t1, LI-t2, ST, IN Non-tumor (69): LI-n, PR, KD, SP
2	16	2	Non-tumor (15): KD Tumor (1): LI-t1
3	7	7	Lymph nodes (3): 12R, 12L Non-tumor (4): SP
4	5	2	Tumor (2): ST Lymph nodes (3): Ax-l
5	4	1	Tumor: LI-t1
6	4	2	Tumor (1): LI-t2 Non-tumor (3): SP
7	3	1	Non-tumor (3): SP
8	3	2	Non-tumor (3): PR, LI-n
9	3	1	Tumor (3): LI-t1
10	3	3	Tumor (3): ST, AX-l, LI-t1
11	3	1	Tumor (3): LI-t1
12 (other)	26	10	Lymph nodes (9): AX-r, AX-l Tumor (12): LI-t1, LI-t2, ST, IN Non-tumor (5): LI-n, PR, KD, SP

AX-r = right axillary lymph node, AX-l = left axillary lymph node, LI-t1 = liver tumor site 1, LI-t2 = liver tumor site 2, ST = stomach, IN = intestine, LI-n = normal liver, PR = prostate, KD = kidney, SP = spleen.

Table 5

Sequence groups for patient IV.

Group #	# of sequences	# of tissues	Tissues
1	123	10	Lymph nodes (73): LN-p, AX-l, LN-o, AX-r, LN-l Tumor (50): KD, SP, DI, IN
2	17	1	Non-tumor (17): ST
3	11	1	Tumor (1): IN
4	4	3	Lymph nodes (4): 12P, 12A, 12L
5	4	1	Lymph nodes (2): 12P, 12o Tumor (2): SP
6 (other)	24	11	Lymph nodes (10): LN-p, AX-l, LN-o, LN-l Tumor (10): 27T, 12A, 20, 27D, 8 Non-tumor (5): ST

AX-r = right axillary lymph node, AX-l = left axillary lymph node, LN-p = periaortic lymph node, LN-o = occipital lymph node, LN-l = lung lymph node, KD = kidney, SP = spleen.

Table 6

Sequence groups for patient GA.

Group #	# of sequences	# of tissues	Tissues
1	86	4	Brain (25): MG, FL Other (61): SP, LN
2	36	2	Brain (36): MG, FL
3	31	3	Brain (12): FL, TL Other (19): LN
4	9	1	Brain (9): TL
5	10	2	Brain (8): MG Other (2): LN
6	10	1	Brain (10): TL
7	6	1	Brain (6): FL
8	5	1	Other (5): LN
9 (other)	35	5	Brain (20): MG, FL, TL Other (15): SP, LN

FL = frontal lobe, MG = meninges, TL = temporal lobe, LN = lymph node, SP = spleen.

Table 7

Sequence groups for patient CX.

Group #	# of sequences	# of tissues	Tissues
1	107	8	Brain (95): MG, BG, OW, OJ, TL, PV, CP Non-brain (12): CO
2	33	5	Brain (20): MG, OW, OJ, CP Non-brain (13): LN
3	8	3	Brain (8): BG, OW, PV
4	7	1	Non-brain (7): CO
5	6	1	Non-brain (6): LN
6	4	2	Brain (4): BG, OW
7	3	1	Brain (3): BG
8	3	1	Brain (3): BG
9	3	3	Brain: (3): BG, TL, OW
10	3	1	Non-brain (3): CO
11 (other)	31		Brain (25): MG, BG, OW, OJ, TL, PV, CO Non-brain (6): CO, LN
12 (not shown)	2	1	Brain: 2 sequences with large deletions

MG = meninges, TL = temporal lobe, BG = basal ganglia, OW, occipital white matter, OJ = occipital junction, PV = periventricular space, CP = choroid plexus, LN = lymph node, CO = colon.

Table 8

Sequence groups for patient AZ.

Group #	# of sequences	# of tissues	Tissues
1	53	5	Brain (21): FL, MG Non-brain (32): LI, LN, SP
2	6	2	Non-brain (6): LN, SP
3	4	1	Brain (4): MG
4	4	1	Non-brain (4): LI
5 (other)	19	5	Brain (5): MG, FL Non-brain: LI, LN, SP

FL = frontal lobe, MG = meninges, LI = liver, LN = lymph node, SP = spleen.

Table 9

Sequence groups for patient BW.

Group #	# of sequences	# of tissues	Tissues
1	93	5	Brain (82): FL-GM, TL, MG, BG Non-brain (11): SP
2	16	1	Brain (16): FL-WM
3	16	1	Non-brain (16): SP
4	7	2	Brain (6): BG, MG Non-brain (1): SP
5	6	2	Brain (4): TL Non-brain (2): SP
6	3	1	Brain (3): MG
7	3	2	Brain (1): MG Non-brain (2): SP
8	3	1	Brain (3): FL-GM
9 (other)	18	6	Brain (12): FL-GM, FL-WM, TL, BG, MG Non-brain (6): SP

FL-GM = frontal lobe grey matter, FL-WM = frontal lobe white matter, MG = meninges, BG = basal ganglia, TL = temporal lobe, SP = spleen.

Table 10

Sequence groups for patient DY.

Group #	# of sequences	# of tissues	Tissues
1	90	8	Brain (78): FL-w, MG, TL, BG, FL-g Non-brain (12): LN, SP, LI
2	15	3	Brain (2): MG Non-brain (13): LN, SP
3	9	1	Non-brain (9): LI
4	7	1	Brain (7): FL-g
5	5	1	Non-brain (5) LN
6	4	1	Non-brain (4): SP
7	3	2	Non-brain (3): LN, SP
8 (other)	35	8	Brain (23): MG, FL-w, TL, BG, FL-g Non-brain (12): LI, LN, SP

MG = meninges, TL = temporal lobe, BG = basal ganglia, FL-w = frontal lobe white matter, FL-g = frontal lobe grey matter, SP = spleen, LI = liver, LN = lymph node.