UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Functional characterization of putative miRNA-like sequences encoded in the HIV genome

Carolina dos Santos Ruivinho

Mestrado em Biologia Humana e Ambiente

Dissertação orientada por:
Professora Doutora Margarida Gama Carvalho

2022

Acknowledgments

First of all, I would like to thank my supervisor Margarida Gama Carvalho for believing in me to take this project forward and for constantly contributing with her guidance, support, and knowledge. It was not an easy project to unpack, but thanks to excellent guidance I never lacked motivation. Thank you for being an example to follow and for helping me to grow immensely as a researcher. I hope I was up to the challenge launched a year ago and that it won't be the last we face together.

To BioISI, thank you for having me in your facilities, promoting quality science, and awarding me with one of the BioISI Junior Programme Research fellowships.

To my colleagues at the RNA Systems Biology Group, no words will be enough. Thank you to everyone who contributed to my learning. A special thanks to Cláudia, Tânia, Lara, and Alexandre, for being a constant presence during this year and making me feel that the lab was a second home. Thank you for your support, I may not finish the thesis knowing how to enter the room smoothly or to close a packet of cookies, but I will finish it with a full heart for having had the privilege of meeting you. An extra special word of affection to Tânia and Cláudia, for never hesitating to take time from their work to contribute to mine, but above all for being the greatest example of resilience and dedication that I could have found in this group, I will take all the teachings you left me to the next steps of my journey.

To those I call family, I deeply appreciate all the support you have given me and all the tools you have placed in my trust for me to build my own path. Thanks to Clara and Pedro for welcoming me as a sister and a special thanks to my godparents for being the biggest support I could wish for. Thank you for encouraging me to be better and to give more of myself despite my stubbornness. To my younger brother, I hope that my path will be an inspiration to you because without a doubt you are my biggest motivation.

To my friends from Coimbra, especially Rita, Diana, and Bea, the biggest thank you in the world. You put up with my dramas, listened to my complaints, helped me find solutions to problems that weren't even yours, and all without complaining (I mean, almost). I thank all the coffees, all the video calls that kept us together, and all the adventures that always made me forget there was work to be done. It will be a pleasure to be able to continue to have you in my journey and to be part of yours.

To those who put up with me in Lisbon, thank you for making this city a little more welcoming and forcing me to take breaks and relax. To Adriana, thank you for all the support, for all the beers, and for celebrating my achievements as much or more than I do. To my housemates, I know that I didn't always have the time or patience for all your plans, but living with you was without a doubt one of the highlights of my life. Thank you, João, André, Ana, Mariana, Diogo, and Jorge for making this year easier, even though you still think that my thesis is about bacteria. A special thanks to Ana for having welcomed me with open arms, for being my football companion, but above all for being a fantastic friend that I will take forward in life.

To my girlfriend, Rita, you know that I appreciate every second by your side and all the motivation you have given me. These last few months have been full of stress and work, but with you by my side, they felt light as a feather. Thank you for all the love, for the constant motivation, and for always being there at the end of the day to make everything better.

“Science and everyday life cannot and should not be separated” – Rosalind Franklin

Resumo

Os microRNAs (miRNAs), também conhecidos por miRs, são um grupo de pequenos RNAs não codificantes envolvidos no controle de uma ampla gama de atividades celulares, como desenvolvimento, função imune e morte celular. Normalmente exibem entre 19 a 22 nucleótidos de comprimento e, em contraste com genes codificantes, apresentam um nível elevado de conservação entre espécies. No processo de biossíntese canónica dos miRNAs, estes são transcritos pela RNA polimerase II dando origem a um transcrito primário (pri-miRNA). Os pri-miRNAs são processados pela DROSHA para dar origem a um precursor com cerca de 70 nucleótidos conhecido como pre-miRNA, que vai posteriormente ser transportado para o citoplasma. Já no citoplasma os pre-miRs são então processados pela DICER levando à produção do miRNA maturo com cerca de 22 nucleótidos que vai acumular-se na Argonauta para formar o complexo de silenciamento induzido por RNA (RISC).

Para além dos miRNAs produzidos pelo genoma do hospedeiro, foram já identificados diversos vírus com a capacidade de originar miRNAs virais (v-miRNAs). A maioria dos v-miRNAs identificados até à data são, no entanto, produzidos em vírus de DNA, na sua maioria pertencentes à família Herpesviridae. Em contraste, a existência de miRNAs codificados por genomas de vírus de RNA tem sido controversa. Por um lado, porque a maioria dos vírus de RNA replica-se no citoplasma, portanto não poderiam aceder à maquinaria de produção de miRNAs canónicos que se localiza no núcleo. Outra razão é que a produção destas moléculas através de um mecanismo de síntese canónico resultaria na clivagem contraproducente do genoma viral e dos seus transcritos. Apesar disto, foi demonstrado que alguns vírus, incluindo retrovírus, podem produzir miRNAs, seja por uma via não canónica, dependente de RNA Polimerase III, como acontece no vírus da leucemia bovina, ou pela via canónica, mecanismo usado pelo vírus da leucose aviária.

O vírus da imunodeficiência humana (VIH) é um vírus de RNA do género lentivírus que se pode apresentar na forma de dois subtipos (VIH-1 e VIH-2). Tal como acontece para outros vírus, a produção de v-miRNAs por parte do genoma do HIV poderia constituir uma vantagem para a sua replicação nas células hospedeiras. A noção de que o HIV-1 poderia dar origem a estas moléculas começou a ser debatida após a identificação de pequenos RNAs no genoma viral que apresentam uma estrutura de stem-loop similar aos miRNAs celulares. Atualmente já se sabe que todas as cinco regiões do genoma do HIV (Tar, Gag, Pol e Nef e 3’ LTR) são prováveis hotspots de produção de miRNAs e identificaram-se 5 potenciais v-miRNAs, 4 deles já catalogados em bases de dados de miRNAs. Contudo, alguns estudos baseados em técnicas de sequenciação de próxima geração (NGS) indicam não conseguir detetar as moléculas previamente identificadas em células infetadas com HIV-1, o que reacende o debate sobre a capacidade do HIV-1 gerar miRNAs. A possibilidade do HIV-2 produzir miRNAs não foi até hoje alvo de qualquer estudo. A descoberta e triagem de miRNAs derivados do genoma do HIV em células infetadas apresenta alguns desafios, uma vez que é aconselhável usar condições experimentais próximas à infeção *in vivo*, bem como populações homogéneas de células-alvo do HIV para limitar a variabilidade na resposta à infeção. Em segundo lugar, é necessário ter em atenção que os v-miRNAs podem apresentar estrutura, comprimento e processos de biossíntese diferentes dos vistos para os miRNAs do hospedeiro. Finalmente, foi já reportado a baixa abundância destas moléculas (aproximadamente 1%) em comparação com miRNAs derivados do hospedeiro, o que implica maiores dificuldades na deteção destas sequências.

Resultados preliminares de um trabalho anterior do nosso grupo, que usou o perfil de espécies de RNA com entre 15 a 30 nucleótidos de comprimento de uma população altamente homogénea de células T naïve CD4+ humanas infetadas por HIV-1 e HIV-2, sugerem que tanto o HIV-1 como o HIV-2 podem potencialmente gerar moléculas semelhantes a miRNA, com evidência de processamento discreto na extremidade 5' e potencialmente decorrentes de hairpins de RNA conservados com cerca de 70 nucleótidos. Com este projeto pretendemos validar as três moléculas identificadas anteriormente como

possíveis miRNAs, duas codificadas pelo HIV-1 (sncHIV1_1 e sncHIV1_2) e uma pelo HIV-2 (sncHIV2_1), como miRNAs genuínos, dissecar o mecanismo usado pelo HIV para produzir miRNAs e investigar o impacto funcional destas moléculas na expressão génica tanto do vírus como do hospedeiro. Os nossos resultados mostram que tanto o HIV-1 quanto o HIV-2 podem codificar fortes candidatos a miRNA que requerem uma estrutura secundária em stem-loop para serem expressos, uma vez que após a inserção de mutações que destabilizam a estrutura secundária, observamos uma diminuição nos níveis de miRNA maduro em duas das três moléculas, uma proveniente de cada genoma viral. Além disso, o knock-down da Dicer levou igualmente a uma diminuição na acumulação de miRNAs maduro correspondentes ao sncHIV1_2 e sncHIV2_1, mostrando que pelo menos dois dos nossos candidatos são derivados de um processamento dependente da Dicer, ao invés de serem meramente produtos de degradação de RNA.

Adicionalmente, usando diferentes algoritmos, previmos ainda que os três miRNAs candidatos possuem diversos alvos no genoma viral e do hospedeiro. Curiosamente uma grande percentagem dos alvos previstos pertencem ao interactoma do HIV-1, o que nos permitiu retirar algumas conclusões relativamente ao modo de atuação dos miRs candidatos e a sua interação com os alvos previstos e as proteínas virais. De um modo geral, esta análise preliminar revelou que há uma prevalência de alvos previstos para estas moléculas semelhantes a miRNAs codificadas pelo HIV-1 com impacto negativo em proteínas virais específicas (p.e Tat), o que sustenta a hipótese de que a expressão dos miRNAs candidatos poderia promover a infeção através da inibição de proteínas do hospedeiro que afetam negativamente as proteínas virais. Além disso, a existência de alvos compartilhados entre os três miRs candidatos e outros miRNAs virais reforça a ideia de que nossas moléculas podem desempenhar um papel na replicação e manutenção do ciclo viral. Adicionalmente, a lista de alvos previstos no genoma do hospedeiro para as três moléculas parece estar enriquecida para importantes processos biológicos, funções moleculares e compartimentos celulares relacionados com a regulação da infeção viral, latência e sobrevivência celular.

Em síntese, os resultados aqui descritos foram capazes de apoiar a validação da existência de miRNAs codificados pelo HIV-1 e HIV-2, que parecem fazer uso dos processos canónicos de biossíntese de miRNAs. Futuros estudos seriam necessários para explorar a participação de outros passos chaves do processamento de miRNAs, como por exemplo o processamento do pri-miRNA pela Drosha, o transporte para o citoplasma pela exportina-5 e a acumulação no RISC. Este trabalho permitiu ainda fornecer uma análise preliminar, mas detalhada dos mecanismos de interação dos candidatos miRNAs na relação vírus-hospedeiro e, desta forma, abrir portas para a dissecação dos mecanismos da patobiologia do HIV-1 e do HIV-2 associados a v-miRNAs.

Palavras-chave: microRNAs; v-miRNAs; HIV-1; HIV-2; interação vírus-hospedeiro.

Abstract

MicroRNAs (miRNAs) are small non-coding RNAs with approximately 19–24 nucleotides long, that play a critical role in the regulation of many viral and host protein-coding genes. These small RNAs are encoded by cellular or viral genomes, however, although DNA viruses have been found to produce miRNAs of similar size to eukaryotic miRNAs, it has been generally assumed that RNA viruses do not encode miRNAs, based on the notion that the generation of canonical miRNAs from a pre-miRNA hairpin would result in useless cleavage of the viral genome and transcripts.

HIV retroviruses (HIV-1 and HIV-2) are RNA viruses that have been intensively studied since HIV-1 infections constitute one of the biggest health issues of the last decades. Nevertheless, the existence of HIV-1 encoded miRNAs is still up for debate, while the existence of HIV-2 encoded miRNAs has never been investigated.

Previous work from our lab using a high-quality dataset of small RNAs from naïve CD4+ T cells infected with either HIV-1 or HIV-2 detected three candidate molecules, two encoded by the former and one from the later viral genome that display typical features of authentic miRNAs. With this project, we aim to experimentally validate the putative HIV-1 and HIV-2 encoded miRNAs as *bonafide* miRNAs, dissect the mechanism used by HIV to produce them, and investigate the functional impact of these molecules in host and viral gene expression. Our results show that both HIV-1 and HIV-2 can encode strong miRNA candidates that require a hairpin secondary structure to be expressed. Moreover, the knock-down of Dicer lead to a decrease in the accumulation of mature putative miRNAs for two of our three candidates, suggesting that they are derived from Dicer-dependent processing, rather than being merely RNA degradation products. Regarding the functional impact of these miRNAs, several viral and host targets were predicted. Importantly, the list of predicted targets for all three molecules seems to be enriched in biological processes, molecular functions, and cellular compartments related to the regulation of viral infection, latency, and cell survival.

Overall, our results were able to support the existence of putative HIV-encoded miRNAs and provide new insights regarding the v-miRNA-dependent regulatory mechanisms linked to the pathobiology of HIV-1 and HIV-2, warranting future in-depth studies.

Keywords: microRNAs; v-miRNAs; HIV-1; HIV-2; viral-host interaction.

Table of Contents

Figures Index

Table Index

Abbreviations

3’LTR	3’ Long terminal repeat
3’UTR	3’ Untranslated region
5’LTR	5’ Long terminal repeat
5’UTR	5’ Untranslated region
AGO	Argonaute
ALV	avian leukosis virus
BFV	Bovine foamy virus
BLV	Bovine leukemia virus
cDNA	complementary DNA
DGCR8	DiGeorge syndrome critical region 8
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
Env	Envelope glycoprotein
Gag	Group-specific antigen
HCMV	Human cytomegalovirus
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HSV	Herpes simplex virus
KSHV	Kaposi's sarcoma-associated herpesvirus
MCMV	Mouse cytomegalovirus
miRs	miRNAs
mRNAs	Messenger RNAs
Nef	Negative factor
NGS	Next-Generation Sequencing
NK	Natural killer
Pol II/III	RNA polymerase II/III
Pre-miRNAs	Precursor miRNAs

Pri-miRNAs	Primary miRNAs
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RT	Reverse transcriptase
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
SFV	Simian foamy virus
siRNAs	Small interfering RNAs
SIV	Simian immunodeficiency virus
sncRNAs	Small noncoding RNAs
TAR	trans-activation response element
TCR	T cell receptor
v-miRNAs	Viral miRNAs

1. INTRODUCTION

1.1. Discovery, biogenesis, and functions of miRNAs

MicroRNAs (miRNAs) are a group of small non-coding RNAs (sncRNAs) that typically display 19-22 nucleotides in length. These molecules act as fine-tuners of gene regulation, ultimately contributing to cell fate and survival, so the discovery of miRNAs turns out to be a note-worthy breakthrough[1].

In 1993 the first miRNA was identified in *Caenorhabditis elegans*[2]. After that, other similar regulatory RNAs were discovered in different organisms[1]. These sncRNAs have been estimated to regulate at the post-transcriptional level approximately 60% of mammalian genes[3], with a single miRNA having the ability to regulate hundreds of mRNAs[1]. In comparison to protein-coding sequences, miRNAs display a higher level of cross-species conservation, indicating that they have been positively selected based on their regulatory function[4]. Various regulatory pathways such as metabolism, apoptosis, proliferation, differentiation, cancer, and so on, have been shown to be severely affected by miRNAs[3].

miRNAs are transcribed as part of a hairpin structure within a larger transcript — the primary miRNA (pri-miRNA). Inside the nucleus, the pri-miRNA is processed to produce the precursor miRNA (pre-miRNA), which is subsequently exported to the cytoplasm and cleaved into a miRNA duplex. Helicase activity results in the unwinding of the duplex into a passenger strand and a functional mature miRNA. The first one is degraded, while the functional strand is put into a RISC (RNA-induced silencing complex) containing Argonaute (AGO) proteins. The RISC is guided to its target transcripts by the functional strand, which has a seed region (nucleotides 2–7) complementary to a region on the target mRNA, usually in the 3'UTR[1]. Figure 1.1 shows the biogenesis of miRNAs and the main proteins involved in the process.

1.1.1. Impact of cellular miRNAs in viral infection

The Human Immunodeficiency Virus (HIV) is an RNA virus belonging to the genera of lentiviruses, a family of retrovirus known for their chronic and persistent infections. HIV is further divided into two categories: HIV-1 (more virulent) and HIV-2. The two subtypes have many of the same viral transmission, replication, and pathogenic pathways, however, they differ in their behavior inside the host cells and their genetic profile[5].

HIV-1 infection is more severe than HIV-2 and is responsible for a global pandemic since an ancestral form of the virus, derived from a simian immunodeficiency virus (SIV)-infected chimpanzee, entered the human population over a century ago[6]. HIV-1 primarily targets and replicates in T lymphocytes expressing the cluster of differentiation 4 surface marker (CD4) and host infection progresses through three main stages: acute, chronic, and AIDS (acquired immunodeficiency syndrome). During the acute infection phase, replication occurs very rapidly, triggering multiple cell death pathways and ultimately progressing to systemic CD4+ T cell depletion, thereby impairing cell-mediated immunity and leading to a systemic failure of the immune system with life-threatening consequences[6]. The control and eradication of the ongoing HIV pandemic, that was already claimed over 35 million lives, remains a difficult challenge[6].

Host microRNAs can influence viral infections, including HIV infection, positively or negatively in a multitude of ways by targeting host factors involved in viral replication (Figure1.2)[6]. These may include regulation of (1) the pathogenicity and viral cycle (2) the efficiency of host innate and adaptive immune

response, and (3) inflammatory response magnitude[7]. Because of this, miRNAs are emerging as both distinct diagnostic indicators and potential targets for the development of new anti-viral therapies[8].

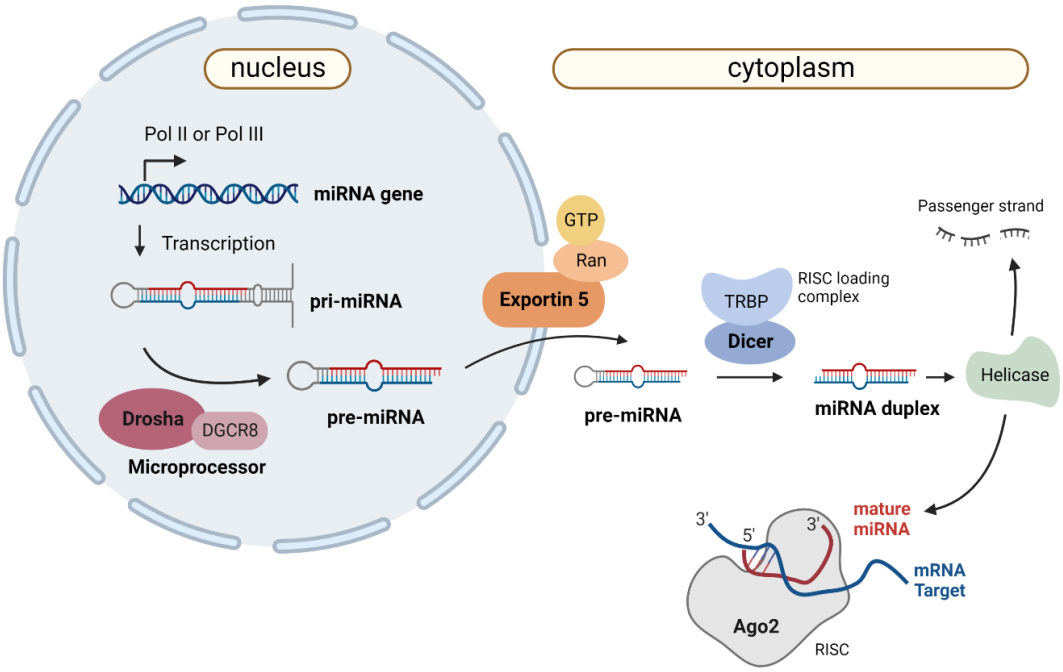


Figure 1.1. Canonical miRNA biogenesis and mechanism of action.
Canonical miRNA biogenesis starts with RNA polymerase II or III (Pol II/Pol III) activity, which transcribes the primary miRNA (pri-miRNA) as a several-hundred-nucleotide-long transcript with a 33-base pair hairpin loop. The microprocessor complex (composed of Drosha and DGCR8) cleaves the pri-miRNA to produce a 60-nt pre-miRNA, which will be transported to the cytoplasm in an Exportin5/RanGTP-dependent manner. The RISC loading complex cleaves the pre-miRNA into a 22-nt miRNA duplex in the cytoplasm. Helicase activity unwinds the duplex into a passenger strand, which is later degraded, and a mature miRNA that is loaded into the AGO family of proteins to form a miRNA-induced silencing complex (RISC). Created with BioRender.com

Cellular miRNAs can positively impact virus replication directly by associating with the virus or viral components to counter the antiviral pathways. A remarkable example was the demonstration that cellular, liver-specific miR-122 binds directly to 5'UTR of the hepatitis C virus (HCV) RNA without affecting the viral mRNA stability but promoting HCV replication[6]. The same phenomenon was not reported for HIV-1 or HIV-2 yet.

MicroRNAs can also indirectly promote HIV replication. Several miRNAs seem to play a role in CD4+ T cell activation, a prerequisite for productive HIV-1 infection. For example, overexpression of miR-132 promoted HIV-1 replication in Jurkat T cells and reactivated HIV-1 in latently infected Jurkat cells[9]. Furthermore, our lab also showed that mir34c-5p leads to changes in the expression of several genes involved in TCR signaling and T-cell activation and modulates the cell environment in a way that promotes viral production[10]. Additionally, some cellular miRNAs have been shown to enable HIV infection by downregulating the levels of antiviral proteins[11]. Cellular miR-34 family and miR-124a downregulate the potassium channel protein TWIK-related acid-sensitive potassium channel 1 (TASK1) and let-7c downregulates the cyclin-dependent kinase inhibitor protein p21, in both cases leading to an increase of HIV-1 replication[11].

On the other hand, miRNAs may also negatively regulate HIV replication. Several cellular miRNAs that impair HIV replication were identified (Figure 1.2). For example, miR-29a has been shown to inhibit the expression of viral gene nef, leading to repressed HIV replication in several studies [12–14]. A

posterior study from Houzet et al.[15] reinforced miR-29 action and identified four new miRNAs that target HIV (miR-133b, miR-138, miR-149, and miR-326)[15].

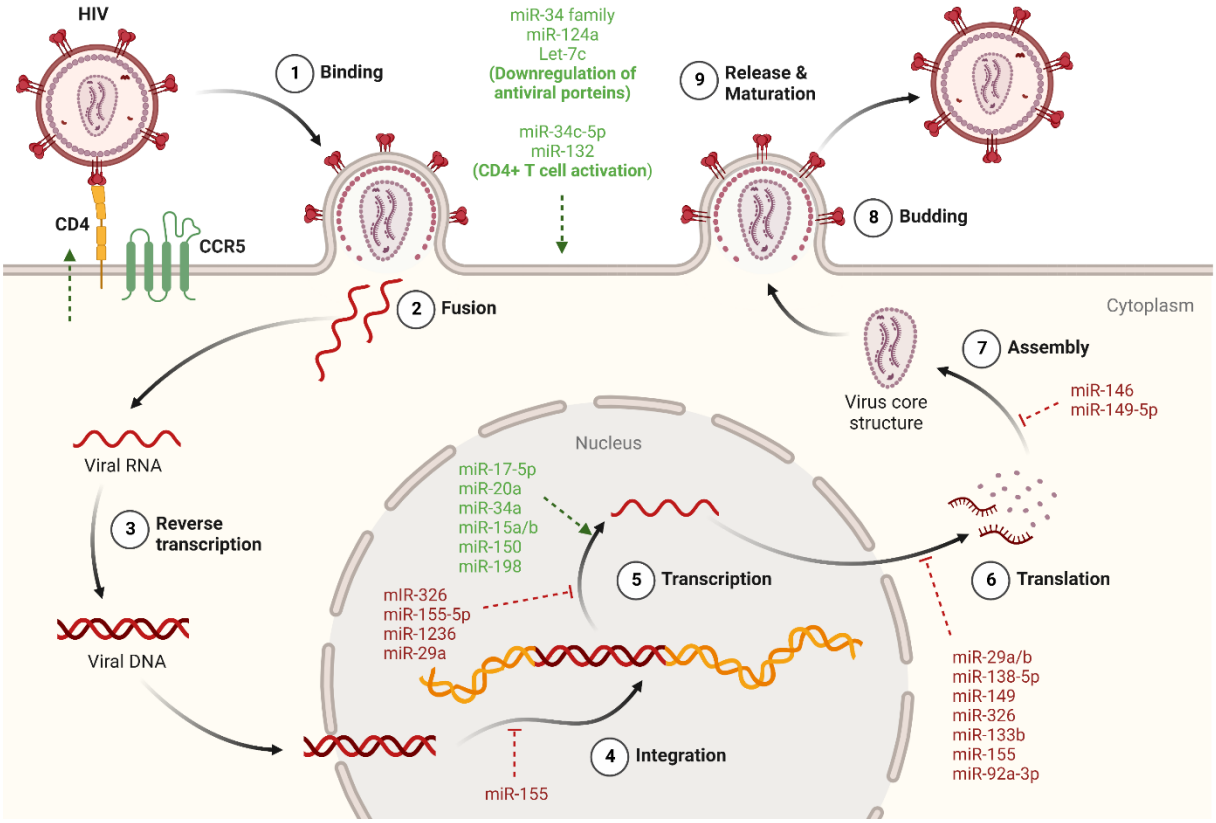


Figure 1.2. Cellular miRNAs modulate HIV-1 infection and replication at various key steps of the viral cycle.
HIV-1 infection can be directly affected by cellular miRNAs that can target HIV-1 viral RNA in a positive (green) or negative (red) manner. While the cellular miRNAs, miR-326, miR155-5p, miR-1236, miR-29a, miR-29b, miR-138-5p, miR-149, miR-326, miR-133b, miR-155, miR-92a-3p, miR-146, and miR-149-5p block viral infection, the cellular miRNAs, miR-17-5p, miR20a, miR-34a, miR15a, miR15b, miR-150, and miR-198 are reported to be promoting the viral replication. In addition, some cellular miRNAs can inhibit host proteins that play an anti-viral role during infection (miR-34 family, miR-124a and Let-7c) or contribute to CD4+ T cells activation (mir-34c-50, and miR-132), promoting by these mechanisms the viral infection. Figure adapted from " Joshi et al., 2022"[16], "Klase et al., 2012"[17], and "Chinniah et al., 2022"[18]. Created with BioRender.com

1.2. Virus-encoded microRNAs

The first viral-encoded miRNA (v-miRNA) was identified in 2004 in the Epstein-barr virus (EBV) through cloning of small RNAs in a cell line infected with EBV[19]. At the moment, VIRmiRNA database contains 1308 entries of experimentally validated miRNAs encoded by 44 viruses[20]. Additionally, the miRbase (Release 22.1), the most used miRNAs database, includes 530 published mature miRNAs entries encoded by 34 viruses[21,22].

The majority of known v-miRNAs are encoded in DNA virus genomes, mainly in the herpesviridae family[19]. Following the EBV-encoded miRNA discovery, several studies identified other DNA viruses able to express miRNAs, such as Kaposi's sarcoma-associated herpesvirus (KSHV) and human cytomegalovirus (HCMV), also members of the herpesvirus family, Papillomavirus, Hepadnavirus (for instance, Hepatite B virus), Adenovirus, and Polyomavirus[19]. The existence of miRNAs encoded by RNA viruses was not widely accepted at the beginning of the study of these molecules. However, there are currently several published studies that show evidence of the existence of miRNAs encoded by RNA viruses, including recent studies regarding SARS-CoV-2[23,24]. In fact, coronavirus-encoded miRNAs were already reported in a work from 2017[23]. However, it was during the recent COVID-19 pandemic that the number of studies on this topic increased dramatically and opened the door to greater acceptance

of miRNAs encoded by RNA viruses. Unfortunately, the number of studies reporting the ability of retroviruses to produce miRNAs is much lower, which makes their acceptance more difficult despite recent advances (discussed in section 1.2.3).

V-miRNAs, similar to host miRNAs, are generated by the Drosha and Dicer machinery (Figure 1.1) and display the typical size range of cellular miRNAs[19,25].Interestingly, unlike eukaryotic miRNAs, these molecules do not seem to display significant sequence conservation across different viruses, rather they have been shown to display function conservation regarding targeted genes and cellular pathways[25,26]. Moreover, v-miRNAs can control the expression of numerous host and/or viral transcripts without triggering an immune response, thus allowing viruses to evade host responses[25].

1.2.1.Functions of viral miRNAs

Viral miRNA regulation is a key component to understand the complexity of viral-host interactions. From the v-miRNAs targets identified so far, it is apparent that viruses-encoded miRNAs play a vital role in immune evasion, viral latency, and regulation of apoptotic processes of the host cells[19,27]. In this context, two main strategies were associated with v-miRNA action: targeting viral gene expression and targeting host cellular processes (Fig 1.3)[19].

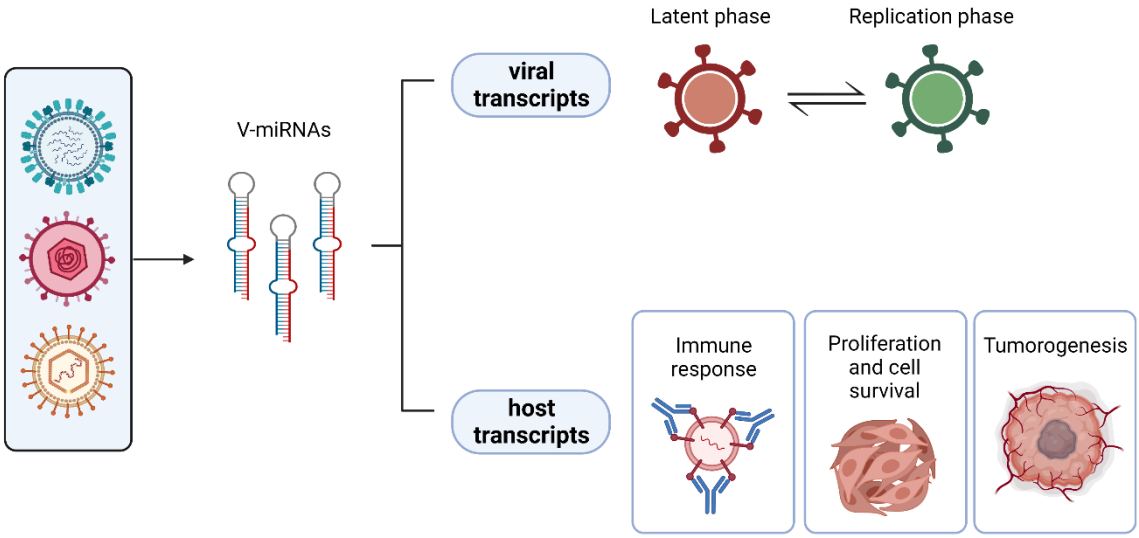


Figure 1.3. v-miRNAs regulate viral and host gene expression
Virally encoded miRNAs can impact viral infection directly by regulating viral genes or indirectly by targeting host genes that play vital roles in infection. The regulation of viral genes by v-miRNAs is normally associated with a mechanism that regulates the balance between the latent phase and the replication phase of the virus. In turn, the targeting of host genes allows the virus to evade anti-viral defenses and prolong the life cycle of the virus through the regulation of different biological processes such as the immune response, apoptosis, and tumorigenesis. Created with BioRender.com

The main reason why viral miRNAs regulate viral transcripts appears to be associated with the promotion of latency or acting as a switch between latency to lytic replication[19]. This mechanism of action is very common in herpesvirus family. HSV-1 encoded miR-H2-3p and miR-H6 promote latency by targeting viral reactivation factors ICP0 and ICP4, which are essential to promote viral reactivation from latency in HSV-1[28]. Similarly, HSV-2 latency is driven by miR-I, miR-II, and miR-III, which downregulate ICP34.5 expression, a key neurovirulence factor[29]. A v-miRNA-dependent strategy is also employed by the human cytomegalovirus, which encodes miR-UL112-1. This v-miRNA regulates lytic replication of HCMV by targeting the IE1 (immediate early viral protein) 3'UTR, thereby promoting latency[30]. Notwithstanding, most viral miRNAs are best known for their role in the regulation of host genes[19].

To date, most studies linking v-miRNAs to the targeting of host cellular genes have been mainly focused on KSHV and EBV infections. miRNAs derived from the KSHV genome target different host genes involved in various cellular processes and pathways such as activation of NF-κB signaling, thus preventing lytic replication[31]. The viral miRNA miR-K12-1 controls cell survival and proliferation by targeting p21, a key tumor suppressor[32]. . Moreover, KSHV miRNAs downregulate C/EBPb p20 (LIP), associated with IL6 and IL10 cytokines regulation[33]. KSHV miR-K1 participates in the regulation of the NF-kB pathway by suppressing IkB[33].

Targeting the host mRNAs may also be a strategy used by viruses to enhance viral survival by silencing cellular genes associated with antiviral immune responses. KSHV miR-K12-11 and miR-K12-5 prevent natural killer (NK) cells from recognizing infected cells[34]. Likewise, HCMV miR-UL112-1, miR-US5-1, and miR-US5-2 regulate mRNAs encoding IL6 and TNF-α, therefore altering host cytokine secretion and ultimately inhibiting the antiviral response[35].

As mentioned above, the ability of RNA viruses to produce miRNAs was not well accepted until recently, therefore, the understanding of their mechanisms of action is not so deeply explored as for miRNAs encoded by DNA viruses. Despite this, the HIV-1 genome was shown to encode proposed pre-miRNAs which, according to computational predictions, target several host cellular genes to create a favorable environment for viral replication[36]. For example, HIV-1 encoded Tar-miR-5p and Tar-miR-3p have been shown to downregulate ERCC1 and IER3, host genes vital for cell survival and apoptosis, thus allowing for HIV-1 infected cells to survive for longer periods by preventing host cell death[37].

These observations indicate that v-miRNAs enhance the viral life cycle, promoting latency, modulating the expression of genes involved in antiviral responses, or regulating cell metabolism. Thus, specific v-miRNAs could be a potential target for antiviral interventions.

1.2.2.Therapeutic Potential of viral miRNAs

The treatment and combat of viral diseases are today one of the greatest health-related challenges. Recent advances in antiviral therapy include specifically inhibiting key components of the viral life cycle in order to constrain viral replication. The identification of v-miRNAs associated with these infections paves the way for the development of new antagomir-based approaches[19]. Antagomirs are a category of chemically synthesized oligonucleotides employed to suppress endogenous microRNAs. Therefore, these oligonucleotides, also known as anti-miRs, can be used against miRNAs that participate in viral infection promotion processes, inhibiting their expression and ultimately reducing the impacts of viral infections on host cells[19]. Several reports support the idea of developing anti-miRs targeting v-miRNAs that participate in viral infection promotion processes rather than host miRNAs, as this reduces the potential side effects and off-target effects and might solve the difficult delivery toxicity-related issues[19].

Such an approach was already applied in several *in vivo* and *in vitro* studies focused on HCV infection. In fact, two studies that developed antagomirs specifically for miR-122, a v-miRNA involved in HCV infection, have entered phase II of human clinical trials and show promising anti-viral effects[38,39]. Additionally, in mouse cytomegalovirus (MCMV) a similar strategy was implemented, and mice receiving anti-miR treatment had a lower incidence of MCMV compared with non-treaded mice[14].

Moreover, v-miRNAs were also considered biomarkers of several viral diseases. A recent study on EBV miRNA BARTT7 showed that this molecule can be an excellent tool to diagnose EBV-associated nasopharyngeal carcinoma and assess treatment effectiveness[40]. Likewise, HCMV- encoded miR-US4-1 is used to predict the efficacy of IFN-α treatment in hepatitis B[41]. Interestingly, the same strategy was used in the fight against Covid-19. SARS-CoV-2 encoded miR-nsp3-3p was monitored in patients progressing from mild/moderate to severe symptoms and was detected in patients' serum at the pre-

severe stage. This v-miRNA serves as a biomarker that helps define priority patients based on their risk of developing severe disease[42].

These promising results suggest that antagomir-based strategies may be strong mechanisms to combat viral infection by inhibiting viral replication and early diagnosing infection-related health complications.

1.2.3. miRNAs encoded by Retrovirus

As previously mentioned, until recently it has been generally assumed that RNA viruses do not encode miRNAs. A possible explanation is that most RNA viruses duplicate in the cytoplasm, therefore the viral RNAs could not interact with the nuclear microprocessor complex and enable the miRNAs biosynthesis[3]. Another reason is the production of canonical miRNAs from a pri-miRNA hairpin would result in unproductive cleavage of the viral genome and its transcripts[3]. However, viruses can regulate cellular pathways out of their self-interest, and in the last decade, several studies have come to contradict the old theory that RNA viruses could not encode miRNAs (most recently reviewed by "Nambo et al."[8]) and the recent COVID-19 pandemic has dramatically increased the study of viral miRNAs encoded by RNA viruses opening doors for a greater acceptance of this phenomenon.

Among RNA viruses, retroviruses are unique because they encode a reverse transcriptase to generate viral cDNA that will ultimately integrate into the host genome, making use of the cell transcriptional machinery to produce a full RNA genomic molecule as well as viral protein-encoding transcripts[43]. Given the nuclear stage in their life-cycle, retroviruses are potential RNA viruses to generate miRNAs.

Several independent groups identified putative miRNAs encoded by retroviruses, including HIV-1, taking advantage of bioinformatic tools. Bovine Leukemia Virus (BLV)[44,45] and Bovine foamy virus (BFV)[46] have been shown to encode miRNAs, that are produced via a Drosha-independent, non-canonical miR-biogenesis pathway. Unlike canonical miRNAs, the BLV and BFV miRNAs are produced by RNA polymerase III (Pol III), which allows a subgenomic transcript of approximately 70 to be cleaved to form a stem-loop hairpin. This harpin is exported to the cytoplasm and subsequently processed by Dicer[25]. According to the VIRmiRNA database, BLV has been shown to encode 9 miRNAs (Table 1.1). One example is BLV-miR-B4-3p, which mimics the seed sequence of host miR-20, a miRNAs associated with B cells tumorigenesis in mice, which may indicate the association of this miRNAs with cell proliferation of infected cells and BLV-induced tumorigenesis[45]. Similar, BFV miRNAs are transcribed by RNA Pol III as a single transcript from the 3'LTR of the viral genome and are subsequently cleaved to generate two pre-miRNAs that are then processed to yield four distinct, biologically active, and experimentally validated miRNAs (miR-bf1-5p, miR-bf1-3p, and miR-bf2-5p, and miR-bf2-3p)[46] (Table 1.1).

On the other hand, the Avian Leucosis Virus (ALV-J) was shown to encode a miRNA produced via the canonical pathway[47]. These observations raise the possibility that retroviral species have the ability to express miRNAs via different biogenesis mechanisms. Based on all publications that describe retrovirus-encoded miRNAs, present and not present in miRNA databases, we prepared a small review of all described retrovirus-encoded miRNAs, their targets, and validation methods, which can be found in Table 1.1. Of note, the molecules studied in this work were not found in the literature as corresponding to any of the HIV-encoded validated and candidates miRNAs identified so far.

1.2.4. HIV-encoded miRNAs

The ability of HIV-1 to produce miRNAs has been the subject of intense debate. The notion that this virus may encode miRNA-like molecules stemmed from the recognition that the HIV-1 genome contains stem-loop structures with the ability to mimic the substrate of the cellular miRNAs biogenesis machinery[6]. Like other miRNAs, the earliest studies regarding HIV-encoded miRNAs relied on

computation approaches. One of the first works on the topic demonstrated that all the five regions of the HIV-1 genome (RNA—trans-activation response (TAR) element, Gag-CA, Gag-Pol frameshift, Nef, and 3'LTR) contain stem-loop structures that are potential miRNA-yielding hotspots[36].

Table 1.1. List of v-microRNAs encoded by Retroviruses family and their proposed targets and functions

Virus	v-miRNAs	Proposed Targets	Proposed Function	Methods	Reference (PMID)
BFV	bfv-miR-bf1-3p/5p	Unknown	Unknown	Northern blot RT-PCR	24522910
	bfv-miR-bf2-3p/5p	Unknown	Unknown	Northern blot RT-PCR	
BLV	blv-miR-b1-3p/5p	Unknown	Unknown	Norther blot Solexa (NGS)	22308400, 23345446
	blv-miR-b2-3p/5p	Unknown	Unknown	Norther blot Solexa (NGS)	
	blv-miR-b3-3p/5p	Unknown	Unknown	Norther blot Solexa (NGS)	
	blv-miR-b4-3p	Mimic of host miR-29	Cell proliferation	Norther blot Solexa (NGS)	
	blv-miR-b5-3p/5p	Unknown	Unknown	Norther blot Solexa (NGS)	
HIV-1	hiv1-miR-h1	AATF, vpr, host miR-140	Apoptosis inhibition, Vpr expression	Cloned	19082544, 20546828
	hiv1-miR-n367	Nef, PABPC4	Viral infection	Northern Blot	15601474, 15722536
	hiv1-miR-tar-3p/5p	Apoptosis-related genes	Apoptosis inhibition and viral propagation	Northern Blot	18299284
	miR-H3*	TATA box in the 5'LTR	Promotes viral replication	Deep sequencing Northern blot RT-PCR	24620741
SFV	sfv-mir-S1	Unknown	Unknown	Norther blot Illumina	24713319
	sfv-mir-S2	Unknown	Unknown	Norther blot Illumina	
	sfv-mir-S3	Unknown	Unknown	Norther blot Illumina	
	sfv-mir-S4	Mimic of host miR-155	Cell proliferation	Norther blot Illumina	
	sfv-mir-S5	Unknown	Unknown	Norther blot Illumina	
	sfv-mir-S6	Mimic of host miR-132	Immunoevasion	Norther blot Illumina	
	sfv-mir-S7	Unknown	Unknown	Norther blot Illumina	
ALV	miRNA-like sequences*	E (XSR) element in the 3′ UTR	Unknown	Northern Blot	24155381

*not present in any miRNA database

The first putative miRNA encoded by the HIV-1 genome to be described and experimentally validated was miR-N367[48]. This molecule was detected through a northern blot using total RNA from MT-4 T cells persistently infected with HIV-1 IIB and SF_2 strains and is derived from a 70-nucleotides-long structure in the Nef/LTR overlapping region. miR-N367 has shown to negatively target the response elements of the LTR U3 region inhibiting the LTR's promoter activity[48,49]. However, this candidate

miRNA was not found in later studies that screened infected HeLa cells[50] and chronically infected T cells using HIV-1 LAV strain[51].

Another reported HIV-encoded miRNA is miR-H1. This miRNA is derived from a stem-loop with 81 nucleotides in length, located downstream of two NF-kB-biding sites in the viral LTR, and has been shown to downregulate the apoptosis antagonizing transcription factor (AATF)[52,53]. Interestingly, this molecule appears to have a high level of sequence variability of both mature miRNA and pre-miRNA across samples of infected patients[54].

Furthermore, two independent studies identified TAT-derived miRNAs. Ouellet et al., reported that miR-TAR-5p and miR-TAR-3p were produced as a result of asymmetrical processing of the HIV-1 TAR element[55]. Later these miRNAs were shown to target host genes related to apoptosis pathways, promoting cell survival and viral replication in HIV-infected cells. More recently, deep sequencing analysis of AGO-1 bound HIV-1 RNAs demonstrated that the 3' region of the TAR hairpin is processed into a miRNA-like molecule[56].

The RT region of the HIV-1 has also been associated with the production of a v-miRNA, named miR-H3 (not present in any miRNA database)[57]. miR-H3 targets the TATA box in the viral 5'LTR promoter and upregulates viral RNA transcription and protein synthesis.

Next-generation sequencing technologies (NGS) are increasingly being used to validate the reported HIV-1 miRNAs and opposite results have been described. A study that relies on sequencing by oligonucleotide ligation and detection (SOLiD) technology-based deep sequencing showed that approximately 1% of the small virus-specific RNAs (mapped to the viral genome) appear to accumulate in mammalian-infected cells. They identified 5 v-miRNA candidates, including the previously described TAR v-miRNA hiv1-miR-n367[58]. In contrast, Whisnant and co-workers[59] generated small-RNA-seq libraries from several cell lines and primary human CD4+ T cells infected with different HIV-1 strains and did not detect the previously described HIV-encoded miRNAs or other novel candidate miRNAs[59]. Furthermore, they reported a bias in the size distribution of the sequences that align to the HIV-1 genome towards 15nts. They concluded that these molecules likely represent degradation products, and that HIV-1 does not encode microRNAs. However, the analysis workflow used by these authors did not take into consideration the possibility that a significant part of these sequences may be host-derived and thus the characteristics of HIV-1-exclusive sequences have not been addressed. Furthermore, a second study also reach the same conclusions as Whisnant when they failed to detect the incorporation of HIV-1-derived sncRNAs or any viral target sequences into the AGO2-RISC complex although they detect a multitude of HIV-1 sncRNAs in HIV-1 infected cells. This suggests that regardless of whether or not HIV-1 can produce miRNAs, they are not involved in the canonical RNAi pathway[60]. The existence of HIV-2-encoded miRNAs has never been addressed.

The discovery and screening of HIV miRNAs in infected cells present several challenges. First, it is advisable to use experimental conditions close to *in vivo* infection, as well as homogeneous populations of HIV target cells to limit the variability of the response to infection. Second, it must be acknowledged that the structure and length of potential miRNA precursors are unknown and may not follow the canonical pathways. Finally, the fact that these molecules may be of extremely low abundance compared to host-derived miRNAs (i.e., less than 1% of all sncRNAs in HIV-1 infected cells as suggested by previous studies[58,60]), implies that very high sequencing depths must be achieved to ensure a reproducible sequence sampling.

Previous work from our lab[10] performed an in-depth RNA-seq study focused on RNA species of between 15-30nt in length to characterize the small RNA profile of a highly homogeneous population of human naive CD4+ T cells and its response to T-cell receptor (TCR) stimulation and both HIV-1 and HIV-2 infection. Of note, this dataset also allows us to inquire about the possible existence of HIV-encoded miRNAs. Therefore, preliminary results of our team, not yet published, used these high-quality datasets to re-assess the presence of HIV-encoded miRNA-like species through the analysis of those

sequences with exclusive homology to the HIV genomes. These results provide evidence that both HIV-1 and HIV-2 can potentially generate miRNA-like molecules, ranging between 19-24 nucleotides, with evidence of discrete 5'end processing, and potentially deriving from conserved RNA stem-loops of ~70nt. Importantly, these HIV-1 and HIV-2 miRNA-like molecules display sequence and RNA structure conservation across HIV subtypes.

1.3. Objectives

With this project, we aimed to experimentally validate the putative HIV-1 and HIV-2 encoded miRNAs previously identified in the lab and investigate the functional impact of the expression of these molecules in host and viral gene expression, focusing on the following specific aims:

1. Obtain experimental validation of candidate miRNAs encoded by the HIV genome by studying a fundamental characteristic that most miRNAs obey: they are generated through Dicer-dependent processing of a hairpin precursor.
2. Identification of potential viral and host targets of novel miRNAs.
3. Study the functional impact of candidate miRNAs in HIV life cycle and generate a regulatory network of virus-host molecular interactions.

To achieve these aims, we integrated distinct approaches focusing on human cell models and bioinformatic tools. Pre-miRNAs previously cloned into an expression vector were transfected into HEK 293 or HEK 293T cells in order to assess their expression under different conditions using RT-qPCR and northern blot. In particular, expression under conditions of hairpin disruption and DICER knock-down were evaluated.

In addition, we applied bioinformatics approaches to predict the targets of candidate miRNAs, unravel their interactions with viral proteins, and evaluate the biological processes in which these molecules may play a relevant role.

2. MATERIALS AND METHODS

2.1. Bacterial Transformation and Sequencing

Transformation and amplification of vector pHTPO containing HIV1_1, HIV1_2, and HIV2_1 wild type and mutant sequences and pcDNA 3.1 – KanFDEST vector containing mir34c-5p was performed using NZY5α Competent Cells (NZYTech, Portugal, Cat. MB00402) according to the manufacturer's instructions. Transformed colonies were amplified using MZYMidiprep Kit (NZYTech, Portugal, Cat. MB05004) according to the manufacturer's instructions, followed by oligos design (Table 2.1.) and sequencing analysis of the results using Eurofins Genomics LIGHTrun sequencing services.

Table 2.1. List of plasmids used in this work, and sequences of primers used for their sequencing.

Plasmid	Fw primer	Rev Primer
pcDNA 3.1 – KanFDEST	CGCAAATGGGCGGTAGGCGTG	TAGAAGGCACAGTCGAGG
pHTPO	CAACGTCGTGACTGGGAAAAC	GAGCGGATAACAATTTCACACAGG

2.2. Cell Culture and Cell Lines

HEK 293 cells were cultured in Minimum Essential Medium (MEM) (Corning, USA, Cat. 10-009-CV) supplemented with 10% Fetal Bovine Serum (FBS) (Biowest, France, Cat. S181BH-500). In the subculture of HEK 293 cells was used Trypsin EDTA 1X (Corning, USA, Cat. 25-053-CI).

HEK 293T Dicer KD cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (Lonza, Switzerland, Cat. BE12-604F) supplemented with 10% FBS (Biowest, France, Cat. S181BH-500) and 2mM GlutaMAX™ Supplement (Gibco, USA, Cat. 35050061). To induce Dicer knock-down , HEK 293T Dicer KD cells were treated with 2 µg/ml of doxycyclin (Dox) (Sigma Aldrich, USA, Cat. D9891-5G). for 48h or 72h.

For transfection cells were seeded at a density of 6x10 cells/ml and transfected with GeneJuice Transfection Reagent (Novagen, Germany, Cat. 70967-3) on the next day with 1 µg of the expression vector.

2.3. RNA and protein extraction.

Total RNA for putative HIVmirs quantification was isolated from HEK 293 cells and HEK 293T cells using either TripleXtractor (GRiSP, Portugal, Cat. GB23.0200), according to the manufacturer's instructions or RNeasy Mini Kit (Qiagen, Germany, Cat. 74104), according to the manufacturer's instructions. RNA concentration was determined using the NanoDropTM1000 Spectrophotometer (Thermo Fisher Scientific, USA).

Total RNA for Dicer qPCR analysis was isolated from HEK 293T using TripleXtractor (GRiSP, Portugal, Cat. GB23.0200) and RNA Clean & Concentrator kit (Zymo Research, USA, Cat. R1019), according to the manufacturer's instructions. RNA was treated with 1U/ml DNase I Set (Zymo Research, USA, Cat. E1010) and concentration was determined using the NanoDropTM1000 Spectrophotometer (Thermo Fisher Scientific, USA).

For western blot, proteins were extracted from the organic fractions of Trizol isolation following an adaptation of Trizol manufacturer's protocol previously reported by Simões et al.[61] .

2.4. cDNA Synthesis and RT-qPCR

To validate the expression of the putative miRNAs, a highly sensitive and specific stem-loop qRT-PCR for the miRNA candidates was performed as described previously[62–64]. Briefly, 40ng of total RNA were

reverse-transcribed with a specific primer for each putative miRNA using NZY M-MuLV Reverse Transcriptase (NZYTech, Portugal, Cat. MB083), according to the manufacturer's instructions. qPCR was performed with Sensifast probe No-ROX (Bioline, USA, Cat. BIO-86005), according to the manufacturer's instructions, and using a Quencher-fluorescent UPL probe #21 (Roche Applied Science) and miR-specific primers. SNORD 48 gene was applied as endogenous control and the plasmid gene H1 was used for normalization of transfection efficiency. All primers sequences are shown in Table 2.2. qRT-PCR product size and specificity were analyzed in 4% Agarose gel.

For Dicer mRNA quantification, cDNA synthesis was done using NZY M-MuLV Reverse Transcriptase (NZYTech, Portugal, Cat. MB083) with random hexamer primers, according to the manufacturer's instructions. qPCR was performed with NZYSpeedy qPCR Green Master Mix (NZYTech, Portugal, Cat. MB224), according to the manufacturer's instructions. GAPDH housekeeping gene was used as the reference gene. Primers for Dicer and GAPDH are shown in Table 2.2. qRT-PCR product size and specificity were analyzed in 2% Agarose gel.

A control with no RT and a no template control (NTC) was performed for each sample. All qPCRs were performed in a BioRad CFX96 Real-Time PCR Detection system.

Relative expression for each gene/miRNA was calculated using the comparative CT method, represented as $2^{-\Delta Ct}$, and plotted using GraphPad Prism software v.9.2.0 (GraphPad, USA).

Table 2.2. Oligonucleotide sequences used in RT-qPCR, and their respective targets.

Target	Primer name	Primer sequence
sncHIV1_1	Stem-loop primer	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACCAGAG CCAACGGTGGCT
	Forward qPCR primer	GCAGGGCAGCATTATCAGAAGG
sncHIV1_2	Stem-loop primer	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACCAGAG CCAACATGCATCG
	Forward qPCR primer	GCAGGGAGTACTGGATGTGGG
sncHIV2_1	Stem-loop primer	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACCAGAG CCAACTGCCTTCA
	Forward qPCR primer	GCAGGGTTGGAACACGGCTGA
mir34c-5p	Stem-loop primer	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACCAGAG CCAACGCAATCAG
	Forward qPCR primer	GCAGGGAGGCAGTGTAGTTAG
SNORD 48	Stem-loop primer	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACCAGAG CCAACGGTCAGAG
	Forward qPCR primer	GCAGGGATGCCATCACCGCAGC
H1	Forward qPCR primer	AAGATGGCTGTGAGGGACAG
	Reverse qPCR primer	TGGTGATTTCCCAGAACACA
Dicer	Forward qPCR primer	TTAACCTTTTGGTGTTTGATGAGTGT
	Reverse qPCR primer	GCGAGGACATGATGGACAATT
GAPDH	Forward qPCR primer	TGCACCACCAACTGCTTAGC
	Reverse qPCR primer	GGCATGGACTGTGGTCATGAG
Universal UPL primer	Stem-loop primer	GTGCAGGGTCCGAGGT

2.5. Dot Blot and Liquid Hybridization analysis

Dot Blot: To test the best avidin (HRP)-conjugated protein dilution and the efficiency of the detection system used in Liquid hybridization a Dot Blot analysis was performed using a biotinylated mimic of mir34c-5p available in the lab. 10pmol of biotinylated mimics (5µM) were loaded directly onto positively charged nylon membranes (GE Healthcare, USA, Cat. RPN203B). Membranes were left drying at room temperature for 15 min and placed on pre-soaked filter papers with 1xTBE buffer and crosslinked using UV crosslinker at 1200µJ for 15 sec. This was followed by baking at 80ºC for 30 min.

After cross-linking and baking membranes were blocked with either 1% Non-fat dry milk (NFDM) blocking buffer in 1x TBST (1xTBS, 0.1% Tween® 20 Detergent) or BSA blocking buffer (NZYTech, Portugal, Cat. MB04602) in 1x TBST for 60min. Blocking was followed by incubation of the membranes with diluted avidin (HRP)-conjugated protein (Bio-Rad, USA, Cat. 170-6528) at room temperature in the blocking buffer. The avidin (HRP)-conjugated protein was used in the following dilutions: 1:300, 1:1000, and 1:2500.

After a 30-min incubation, the membranes were washed in the washing buffer (1× TBS, 0.1% SDS) for 30 min at room temperature and then subjected to signal detection.

For both dot blot and Liquid hybridization, the HRP signals on the hybridized membranes were detected through chemiluminescence detection with Clarity Western ECL Substrate (Bio-Rad, USA, Cat. 1705060) and images were acquired using ChemiDoc XRS+ (Bio-Rad, USA).

Liquid Hybridization: The protocol of Ahmad et al. [65] was used for liquid hybridization based on the Northern blot technique for miRNAs with a few modifications. Briefly, 1µg or 2µg of diluted RNA was mixed with 10pmol of a biotinylated probe specific for each putative miRNA and hybridization buffer up to 25µl of total reaction volume. The following hybridization buffers were tested: Buffer 1 (500 mM Tris-HCl, pH 8.3, 750 mM KCl, 30 mM MgCl2, 100 mM DTT) Buffer 2 (10 mM EDTA, 0.3 M NaCl, 30 mM sodium phosphate buffer at pH 8.0) and Exonuclease I buffer (New England BioLabs, USA, Cat. M0293)

The DNA was denatured at 94 ºC for 10min followed by stepwise annealing for specific periods using C1000 Touch Thermal Cycler (Bio-Rad, USA) as follows: 72 °C for 30 min, 60 °C for 60 min, 55 °C for 30 min, 42 °C for 60 min, 37 °C for 30 min. After completion of the hybridization reaction, the hybridized samples were digested with 1 unit of Exonuclease I (New England BioLabs, USA, Cat. M0293) overnight at 37 °C. The digested mixture was either stored at −20 °C until further use or subjected to electrophoresis for northern blot analysis.

For electrophoresis, a 12% TBE (Tris/Borate/EDTA)-acrylamide gels was used. After polymerization, the gels were pre-run with 1x TBE running buffer for 30 min at 150 V followed by loading of hybridized samples mixed with Tritrack DNA loading dye (Thermo Scientific, USA, Cat. R1161). Then gels were run for 40–90 min 150 V or until the bromophenol dye reached two-thirds of the gel.

After the completion of gel electrophoresis, the acrylamide gels were stained with Green Safe Premium diluted in 1xTBE buffer. This was followed by pre-soaking the gel with 1× TBE. RNA in the gels was transferred onto positively charged nylon membranes (GE Healthcare, USA, Cat. RPN203B) pre-soaked in the same buffer. Electroblotting was performed at 18 mA for 1h. After the completion of the transfer, membranes were placed on pre-soaked filter papers with 1× TBE buffers and crosslinked using UVILink CL 508 (UVITec, UK), at 1200µJ for 15 sec and baked at 80ºC for 1h.

The membranes were washed twice with high stringency buffer (0.1× SSC, 0.1% SDS) for 5min before blocking. After washing, membranes were blocked in a 25 mL BSA blocking buffer (NZYTech, Portugal, Cat. MB04602), in 1x TBST (1xTBS, 0.1% Tween® 20 Detergent) for 60min. Blocking was followed by incubation of the membranes with avidin (HRP)-conjugated protein (diluted 1:2500 with blocking buffer) at room temperature. After a 30-min incubation, the membranes were washed in the

washing buffer (1× TBS, 0.1% SDS) buffer for 30 min at room temperature and then subjected to signal detection.

2.6. Western Blot analysis

For HEK 293T Dicer KD protein extracts 6x SDS-PAGE Sample buffer (31,25 mMTris-HCl, pH 6.8, 12,5% glycerol, 1% SDS, 0.05% bromophenol blue, 10 mM DTT, 200 U/ml benzonase, 0,5 mM MgCl2) was added and samples were heated for 5min at 95ºC. 30 µL of protein sample was loaded into an 8% polyacrylamide gel and resolved at 60V in the stacking gel and 120V in the resolving gel. The proteins were transferred to a PVDF membrane (GE Healthcare, USA, Cat. 10600029) for 60 min at 100V, followed by blocking of the membrane with 5% (w/v) non-fat dry milk, for 1h at room temperature.

For probing was used anti-Dicer1 (1:5000 dilution, clone 5D12.2, Millipore, USA, Cat. 04-721), anti-α-Tubulin (1:2000, clone DM1A, Millipore, USA, Cat. 05-829) at 4ºC, overnight.

After the incubation with the primary antibodies, membranes were washed 3 times in 1xTBST, following incubation with the secondary antibody goat anti-mouse IgG (H+L)-HRP Conjugate (1:3000 dilution, Bio-Rad, USA, Cat. 170-6516). The membranes were washed as previously, and subsequent detection was carried out through chemiluminescence detection with Clarity Western ECL Substrate (Bio-Rad, USA, Cat. 1705060) and images were acquired using ChemiDoc XRS+ (Bio-Rad, USA).

2.7. Statistical Analysis

The statistical tests used are mentioned in the corresponding figure legends and were implemented using GraphPad Prism software v.9.2.0 (GraphPad, USA). Hypergeometric distribution analysis was performed on R software V4.2.1.

2.8. miRNA target prediction

Viral genome binding sites: The two following algorithms were used for the prediction of putative HIV-encoded miRNAs binding sites within the viral genome: Miranda[66] (https://cbio.mskcc.org/miRNA2003/miranda.html), which uses seed match, conservation of the whole target site, and binding energy of the duplex structure to predict miRNAs complementarity to 3'UTR regions, and RNA22 v2[67] (https://cm.jefferson.edu/rna22/), a method for identification of miRNA binding sites and their corresponding heteroduplexes that relay on seed match, free energy, G:U pairs allowed in the seed, sensitivity/specificity, and pattern recognition to predict miRNAs binding sites to a reference genome. It is important to reiterate that multiple algorithms were used since these programs will often predict distinct miRNA binding sites.

For Miranda software, we use the following settings: Gap Open Penalty: -9.000000; Gap Extend Penalty: -4.000000; Score Threshold: 140.000000; Energy Threshold: 1.000000 kcal/mol; Scaling Parameter: 4.000000. The specification used in RNA22 software were sensitivity vs specificity: 63%/61%; Seed region size: 7 nucleotides with the maximum of 1 un-paired base in the seed; Maximum folding energy: -10; Maximum number of G:U: infinite.

The reference genomes for HIV-1 (GCF_000864765.1), and HIV-2 (GCF_000856385.1) are publicly available at NCBI and were used in all our analyses.

Human transcriptome binding sites: Prediction of sncRNAs targets was done using the three following algorithms: TargetScan[68](https://www.targetscan.org/vert_50/seedmatch.html), TargetRank[69] (http://hollywood.mit.edu/targetrank/) and miRDB[70] (http://www.mirdb.org/index.html). All three tools use seed match, conservation, and local AU content. miRDB and Target scan also use free energy and site accessibility for the prediction of putative targets. Moreover, for miRDB, we submitted the entire

sequence of our predicted miRNAs, while the other algorithms used the seed sequence (position 2-8 of the mature miRNA).

Homology with host and viral miRNAs: To search for sequences homologies between HIV-encoded miR-like sequences and Human-derived miRNAs listed in miRBase Release 22.1, we used the search tool from miRBase[21,22,71] available here: https://www.mirbase.org/search-rnacentral.shtml. ´

Moreover, we use the VirmiRNA database[20], which contains information about experimentally validated viral miRNAs and their targets, to search for common targets between our molecules and other viral miRNAs.

2.9. miRNA-mRNA-protein interaction networks

The HIV-1 Human Interaction Database used for the construction of our networks is publicly available at NCBI[72].

Once bioinformatic analyses have been performed we loaded the HIV-1 Interactome into Cytoscape v3.9, using the "HIV-1_Prot_Name" column as source node, "Human_geneSymbol" column as target node, and "Keyword" column as interaction type. We then investigate whether our target lists were present in this network and the predicted targets present in the HIV-1 interactome along with their first neighbors were used to create an individual network for each candidate miRNA constituted by the interactions between the predicted targets, and the viral proteins.

To search for interactions between the viral proteins and our target genes we use the same network for sncHIV1_1 and sncHIV1_2. Only the "upregulated by" or "activated by" edges, for positive interaction, and "downregulated by" or "inhibit by" edges, for negative interactions were selected and the type of edges were manually modified for better visualization.

For all networks, the size of the nodes was attributed according to the number of interactions in the respective network.

2.10. Functional Analysis

To assess the total set of biological processes in which our targets participate we use the function characterization tool available at http://pantherdb.org/.

We further evaluate whether our targets featured enrichment of Go terms (biological processes, molecular function, and cellular component), KEGG pathways, and Panther curated pathways in comparison to the HIV-1 interactome used as background. This analysis was performed and visualized using ShinyGO 0.76[73] (FDR cutoff: 0,05; # pathways to show: 20; Pathway size 2-2000) and the result lists were sorted by the average of the ranks by FDR and Fold Enrichment.

For both analyses, the consensus gene lists of targets for the HIV-encoded miRNA-like molecules predicted by at least two algorithms and a sub-list of these targets present in the HIV-1 interactome were used.

3. RESULTS

3.1. Cloning and Expression analysis of HIV-encoded miRNAs

3.1.1. Description of the vector construction strategy

Previous work done by our group focusing on deep-sequencing of HIV-infected human CD4 T cells provided evidence that both HIV-1 and HIV-2 can potentially generate miRNA-like molecules, which could have an impact on the production of viral proteins as well as host genes involved in immune response pathways. We aimed to assess whether these molecules, two encoded in the HIV-1genome and one in the HIV-2 genome, predicted *in silico* to be contained in a genomic region presenting pre-miRNA-like features, were *bona fide* canonical miRNAs.

To verify if the putative sncRNAs are processed by the miRNA machinery and ultimately understand the functional impact of the novel HIV-encoded miRNAs on HIV replication, we began by transfecting HEK 293 cells with pHTP0 (NZYTech, Portugal, Cat. MB281) vectors containing the putative pre-miRNA sequences immediately downstream of H1 Pol III promoter (vector map in Figure 3.1A).

The promotor + pre-miR sequence in the expression vector pHTP0 was commercially synthesized, validated by sequencing, and already available in the lab at the beginning of this work. In the present work, these vectors containing the pre-miRs or related sequences were used in every experiment to evaluate the expression of mature miR molecules, which can be detected using the stem-loop RT-qPCR method (Figure 3.1B).

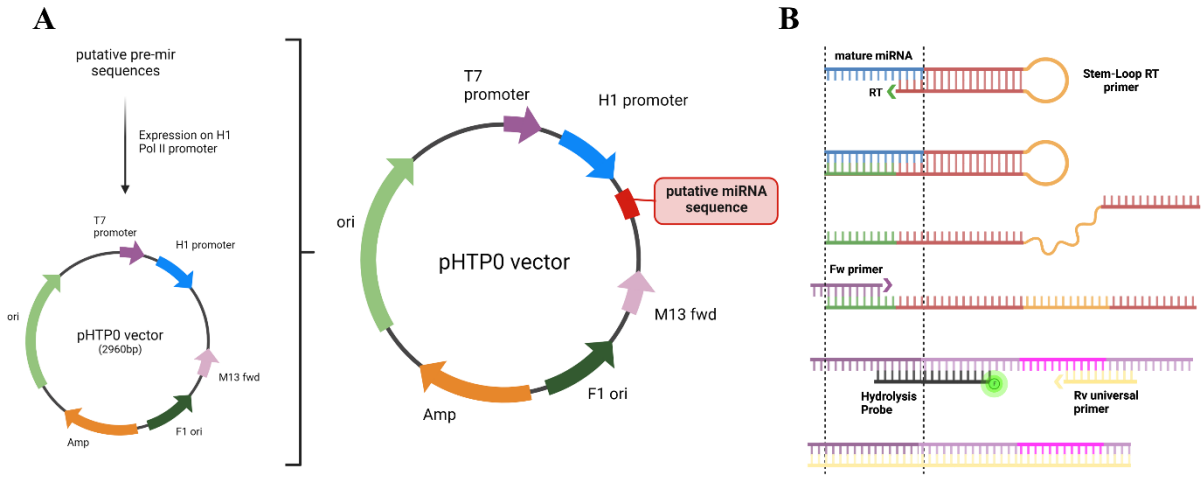


Figure 3.1. Schematic representation of putative miRNAs expression analysis strategy.
(A) A generalized map of the pHTP0 plasmid used for the expression of putative miRNAs and a simplified schematic of the insert of miRNA-like molecules into the expression vector. The putative pre-miRNAs were synthesized and cloned into the pHTPO vector. The cloned vectors were transfected into HEK cells and analyzed by Stem-Loop RT-qPCR (B) Illustration of stem-loop RT-PCR method to amply mature miRNAs adapted from[62]. A stem-loop RT primer is designed to bind the target putative miRNA and contains a universal reverse primer sequence on the loop and a 3' end sequence that are complementary to the 3' end of the mature miRNA. The cDNA is amplified with specific forward and universal reverse primers. The amplification products were detected in this work by real-time PCR using a Quencher-fluorescent probe. Created with BioRender.com

3.1.2. Analysis of putative miRNAs expression using stem-loop qPCR

Because predicted miRNA candidates may be a random degradation product from the viral genome, we carried out a plasmid-derived miRNA expression analysis to exclude this possibility. The predicted pre-miRNAs cloned into the pHTPO expression vector were transfected into HEK 293 cells.

Total RNA isolated from cells transfected with the expression plasmid was subjected to a stem-loop RT-qPCR analysis as previously described[62–64]. Our results revealed that there was no relevant RT-PCR signal in un-transfected HEK 293 cells for any of the sequences, in contrast with what was observed in cells transfected with the expression vectors containing the HIV1_2 and HIV2_1 hairpins (Figure 3.2B). The results shown are normalized for transfection efficiency using the H1 promoter sequence present in the pHTPO plasmid. Although the primers used for transfection normalization target the human RNA Pol III H1 promotor sequence that is present both in the plasmid and genomic DNA, the difference in signal between untransfected and transfected cells is of more than 3 orders of magnitude (>4*10^3, from Ct ~32 to Ct ~ 20 in transfected cells). As such, any signal that might originate from genomic DNA present in the sample is too small to interfere with the normalization, which justifies the use of this gene for transfection normalization purposes in this work. This also shows that we obtained an efficient transfection for our three molecules, including sncHIV1_1, and that, therefore, the reported results are not influenced by transfection problems.

Moreover, gel electrophoresis confirmed the detection of a single PCR product of the expected size for these two constructs (Figure 3.2B/C), but not for HIV1_1, which seems to accumulate low levels of non-specific products (Figure 3.2B).

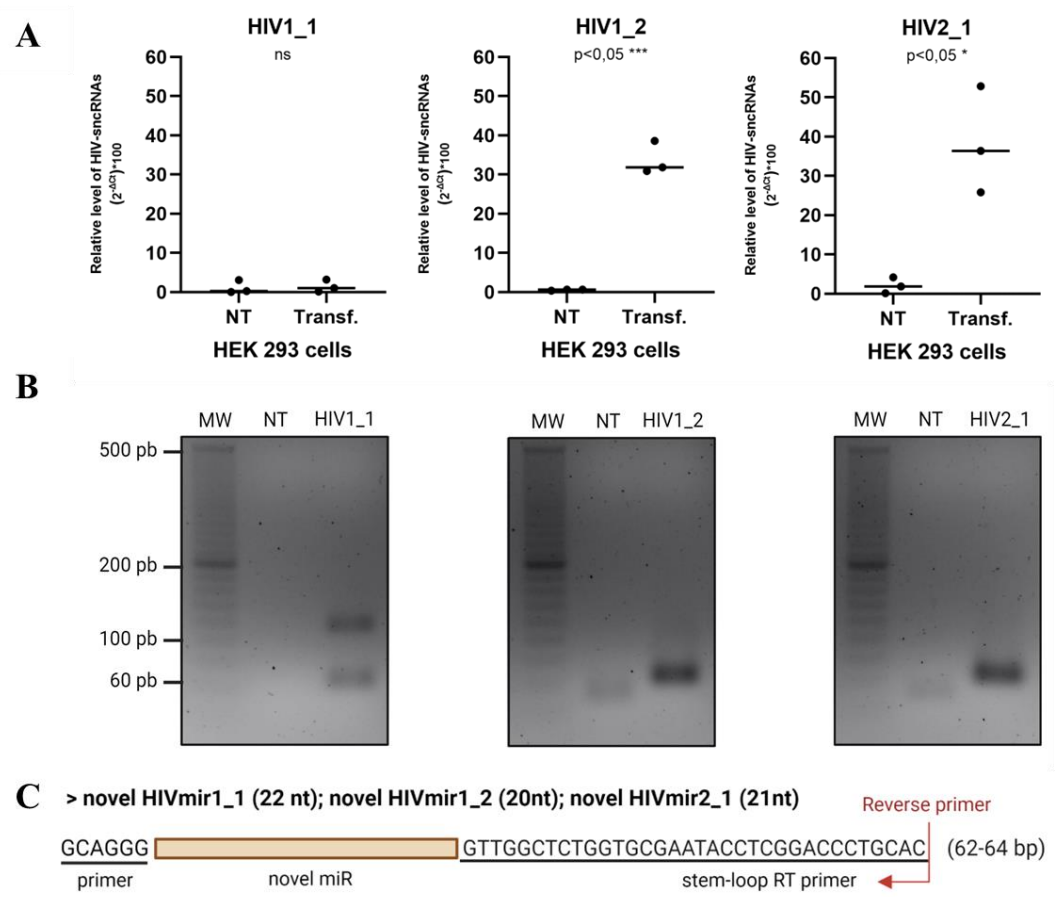


Figure 3.2. Detection of HIV-encoded miRNA-like molecules from plasmid-expressed mature miRNAs.
(A) The relative expression level analysis of sncHIV1_1, sncHIV1_2, and sncHIV2_1 by stem-loop RT-qPCR method 24h after transfection of HEK 293 cells (Transf.) with pHTP0 containing the pre-miRNA sequences of putative miRNAs in contrast with non-transfected HEK 293 cells (NT). SNORD was used as endogenous control and results were normalized to transfection efficiency using the H1 promoter gene present in the pHTP0 vector. Data represent three independent experiments performed in triplicate. *** P<0,0005 *P<0,05 ns= not significant. (B) Analysis of RT-qPCR products for the detection of mature putative miRNAs expression after transfection of HEK 293 cells in contrast with not transfected cells (NT) by Agarose gel 4%. The size of the PCR products was approximately 80bp. MW indicates the use of NZYDNA Ladder IV (NZYTech, Portugal, Cat. MB058). (C) Schematic representation of expected RT-qPCR products.

These findings suggest that viral miRNAs HIV1_2 and HIV2_1 can be expressed in HEK 293 cells via plasmid-derived miRNA expression and indicate that candidate miRNAs are indeed derived from the slicing of predicted precursors rather than random degradation products of viral genome RNAs.

Of note, after having detected predicted HIV-encoded miRs in transfected HEK 293 cells with a qPCR-based method, we tried to implement a liquid hybridization assay based on the northern blot technique as previously described[65] to further confirm the presence of pre- and mature putative viral miRNAs. Regrettably, we did not detect signals of the pre- and mature miRNA (data not shown). The major reason for this result is possibly due to the lack of specificity of the synthesized primers and difficulties in optimizing the reagents used in this procedure, which appear to degrade the RNA.

3.2. Effects of harpin disruption on mature miRNA accumulation

3.2.1. Site-directed mutagenesis method for pHTPO vector containing HIV-encoded miRNAs

Once the validation of candidate miRs expression was accomplished, we next sought to understand whether the formation of the hairpin structure typical of a miRNA was essential for the expression of the putative HIV-encoded miRNAs.

MicroRNAs are processed from hairpin-containing primary transcripts (pri-miRNAs). Pri-miRNAs are generally kilobases in length and have a characteristic internal stem-loop structure containing a double-stranded stem region and an apical loop. Canonically pri-miRNAs are processed in the nucleus by the Microprocessor, which is composed of Drosha and DGCR8 in animal cells, to generate shorter precursors (pre-miRNA) containing a ~22-nt stem and an apical loop of varying length (Figure 3.3A). Pre-miRNAs are subsequently transported to the cytoplasm and cleaved by the Dicer to release a miRNA duplex of a miRNA and a passenger strand (see also Figure 1.1).

Although there are no common sequences conserved among pre-miRNAs, they seem to share several structural features: a 2-nt 3′ overhang, a double-stranded stem region, mismatches in the stem region, and a thermodynamically unstable, single-stranded terminal loop region[74] (Figure 3.3A).

Although the mechanisms by which Dicer recognizes pre-miRNAs remain poorly understood, data suggest that the Dicer recognizes the single-stranded terminal loop structure or pre-miRNAs and selects the sequences based on loop size and length from the 3’ overhang to the terminal loop[74].

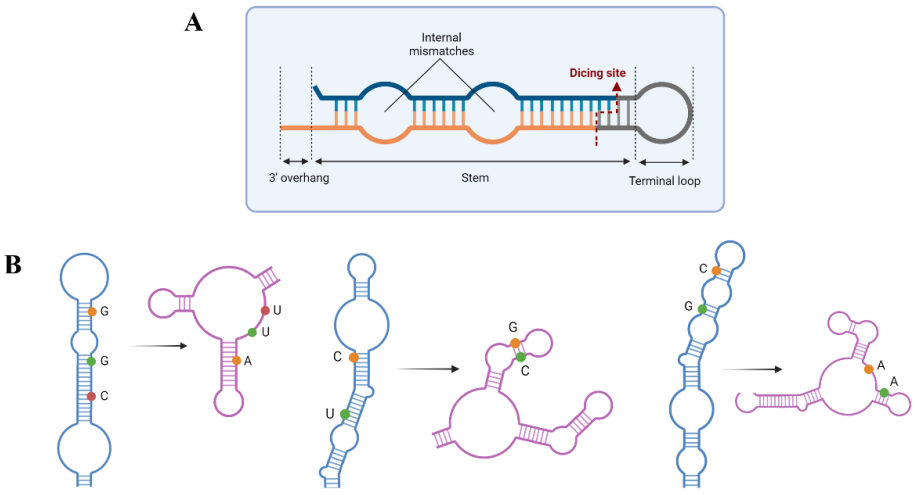


Figure 3.3. Pre-miRNA hairpin disruption using a site-directed mutagenesis protocol
(A) Structural characteristics of typical pre-miRNAs adapted from[74]. (B) Graphic representation of stem-loop mutations. A site-directed mutagenesis protocol was implemented on the wild-type vectors. The result was simulated on the website https://eternagame.org/ where the destruction of the hairpin was demonstrated. Created with BioRender.com

To determine if our putative miRNAs required the conserved stem-loop pre-miRNA formation with these specific characteristics, we took advantage of a previous work from our lab that implemented a site-directed mutagenesis protocol on the wild-type vectors aiming to create mutations able to disrupt the hairpin organization without changing the seed sequence of the candidate miRNA (Schematic representation of these mutations is shown in Figure 3.3B).

3.2.2. Differential expression of Wild-type and Mutant miRNAs in HEK 293 cells

To evaluate the impact of the mentioned mutations we used total RNA samples extracted from HEK 293 cells following transfection of either wild-type or mutant vectors and quantified the accumulation of the mature miR using the stem-loop RT-qPCR method as previously described.

The data of three independent assays for each vector were normalized for transfection efficiency using the H1 gene present in the expression vector and are represented in Figure 3.4. Our results reveal that the mutant sequences lead to a significant decrease, 88%, and 68%, in the accumulation of the mature sncHIV1_2 and sncHIV2_1, respectively, in comparison to the wild-type vectors.

As before, the candidate miR HIV1_1 was only detected at very low levels and no significant difference was observed between mutant and wildtype vectors.

These results suggest that HIV-encoded miRs HIV1_2 and HIV2_1 require hairpin formation for their expression, supporting the hypothesis that both HIV-1 and HIV-2 genomes can express small non-coding RNAs that display the typical features of pre-miRNAs.

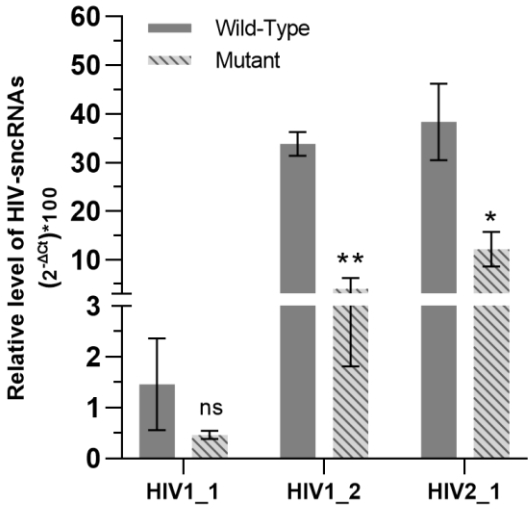


Figure 3.4. Differential expression of Wild-type and Mutant miRNAs in HEK 293 cells.
Expression levels were reduced after introduction of a mutation able to disrupt the hairpin formation. Relative levels of HIV-encoded sncRNAs were evaluated by stem-lopp RT-qPCR 48h after transfection of HEK 293 cells with pHTP0 containing the pre-miRNA sequences of putative miRNAs. SNORD was used as an endogenous control and results were normalized to transfection efficiency using the H1 promoter gene present in the pHTP0 vector. Data represent three independent experiments performed in triplicate. **P<0,005 *P<0,05 ns= not significant (Unpaired Student's t test with Holm-Šídák method). Error bars, means ± SEM.

3.3. Testing DICER-dependent processing of miRNA candidates

Dicer is well known as a key protein in the canonical cellular miRNA processing pathway, being responsible for processing pre-miRNAs into mature miRNA duplexes[1]. It is also well recognized that the reduction of vital components of the miRNA biogenesis pathway results in the attenuation of miRNA steady-state levels[75]. Therefore, it is expected that the depletion of Dicer protein leads to a decrease in the accumulation of canonical mature miRNAs.

In view of the results above we hypothesized that HIV-encoded miRs were a product of the Dicer processing pathway rather than random degradation products of the viral genome. To confirm this assumption, we took advantage of a HEK 293T cell line available in the host laboratory, inducible for shRNA knock-down of DICER. In these cells, the addition of doxycycline (Dox) induces the expression of shRNAs that are subsequently processed into siRNA-like molecules. Although the knock-down

approach has limitations due to incomplete repression of targeted genes, we considered it to be a practical and rapid strategy for determining the importance of Dicer in HIV-encoded miRNAs processing.

Firstly, we carried out a knock-down induction test to ensure that this cell line induction system was functional. HEK 293T cells with inducible shRNA knock-down for Dicer were treated with Doxycycline (1µg/ml) and two time points (48h and 72h) were tested. The silencing efficiency of Dicer mRNA and protein were monitored by RT-qPCR using the Syber Green method and Western Blot analysis, respectively, however, we were unable to detect DICER protein using Western Blot probably due to problems with the available antibody, therefore we assume that the results obtained for the mRNA represent what is happening with the protein.

As shown in Figure 3.5A, Dicer mRNA expression levels were significantly reduced after Dox treatment at both time points in comparison to non-treated control, indicating that after 48h we can obtain satisfactory levels of Dicer depletion (70%) in this system.

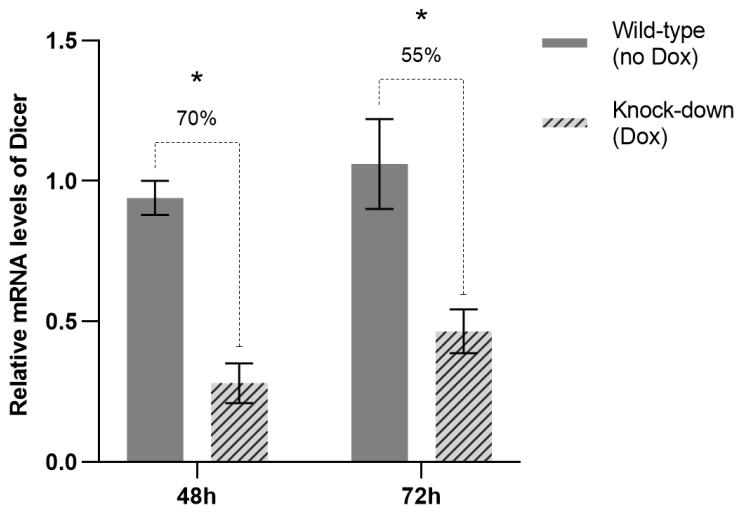


Figure 3.5. siRNA-mediated Knock-down of Dicer reduces the expression of Dicer mRNA levels (A) DICER mRNA levels were reduced in inducible knock-down cell lines. Cells were harvested at 48h and 72h after Dox treatment and mRNA levels of Dicer were analyzed by RT-qPCR using Syber green method. Data represent three independent experiments performed in triplicate. Error bars mean ± SEM. *p< 0,05 (Unpaired Student's t test with Holm-Šídák method). GAPDH was used as an endogenous control.

After validation of our Dicer Knock-down system, we then proceeded in an attempt to understand the influence of Dicer depletion on the processing of our candidate miRNAs.

The differences in viral miRNA abundance between Dicer Knock-down cells and negative control cells (no DOX) were evaluated by RT-qPCR at 48h after Dox treatment. As shown in Figure 3.6, a significant reduction of relative viral miRNA levels was observed for sncHIV1_2 (~59% of decrease) and sncHIV2_1 (~56% of decrease) in Dicer knock-down cells, compared with negative control (P< 0,05). To assess whether the same impact would occur on *bonafide* miRNAs, we used an expression vector containing mir-34c-5p as a control. As seen for our candidate miRNAs, mir34c-5p relative miRNA levels were significantly reduced (~60%) in Dicer knock-down cells compared to the negative control.

Regarding the sncHIV1_1, once again, we were only able to detect low levels of expression, with no significant difference observed between Dicer Knock-down cells in comparison with the negative control cells (Figure 3.6).

These data demonstrated that depletion of Dicer leads to the decrease of putative HIV-encoded miRNAs and suggest that two of our three sncRNAs are a product of the cellular Dicer-dependent processing pathway, rather than random degradation of viral genomes.

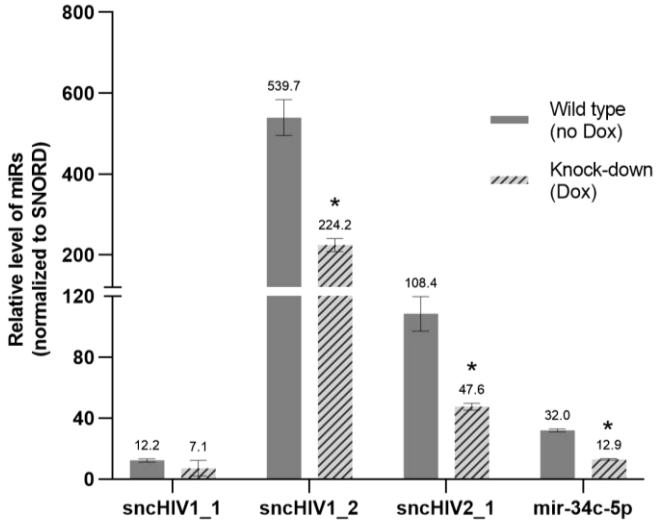


Figure 3.6. . siRNA-mediated Knock-down of Dicer decreases accumulation of mature putative miRNAs.
Relative expression levels of plasmid-expressed sncHIV-encoded molecules and mir34c-5p were significantly reduced in Dicer-deficient HEK293T cells compared with control cells (without knock-down). Significant differences in expression levels were detected by Student's t-test (*P<0,05). Error bars mean ± SEM.

3.4. Characterization of the functional impact of miRNAs encoded in the HIV genome

3.4.1. Bioinformatic prediction of potential targets of novel miRNAs

One of the known biological characteristics of genuine viral miRNAs is that they can modulate the expression of both virally encoded and host genes, regulating processes critical for viral infection[25,76]. Thus, we next aimed to investigate the ability of our HIV mir-like molecules to target functionally relevant mRNAs using an *in silico* approach.

Firstly, we searched for possible targets of our putative HIV-encoded miRs in the corresponding viral genome. Although the exact rules regarding the interaction of viral-encoded miRNAs with their targets are yet to be defined, target prediction tools have three features in common: seed match, conservation, and free energy. We thus selected the following prediction tools, Miranda[77], which searches for seed matching and free energy, and RNA22 v2[67], which identifies miRNA binding sites using a large range of features, such as seed match, free energy, pattern recognition, sensitivity/specificity settings, and G:U pairs allowed in the seed. Since these two algorithms explore different features of miRNA target prediction, the selection of common target sites would thus provide us with more robust evidence of interactions. The predicted target site in the HIV-1 or HIV-2 genome for our three candidate miRs is shown in Table 3.1.

Table 3.1. Potential regulation of HIV-1 or HIV-2 genes by HIV-encoded sncRNAs

HIV-encoded miRNA-like molecule	Algorithm	Target	Position
sncHIV1_1	Miranda	-	-
	RNA22	Env gp120, Tat, Rev	7585-7607
sncHIV1_2	Miranda	Gag	969-991
		Pol	3017-3039; 2478-2500
	RNA22	Gag	969-991; 1161-1181
		Pol	2256-2277; 2478-2500
		Env gp120, Tat, Rev	5868-5891; 6487-6507; 6591-6613; 6609-6630
		Nef	8543-8566
sncHIV2_1	Miranda	-	-
	RNA22	Env	6771-6791; 8298-8323; 9229-9248
		Nef	9229-9248

We noted that RNA22 predicted the modulation of several viral mRNAs, while Miranda predicted only two biding sites, both for HIV1_2, with gag and pol being identified as potential targets for HIV1_2 by both Miranda and RNA22 in the same position (Figure 3.7A). Of note, the binding sites were screened to make sure they did not match the sequence of the miRNAs in the corresponding genome (Figure 3.7A/B).

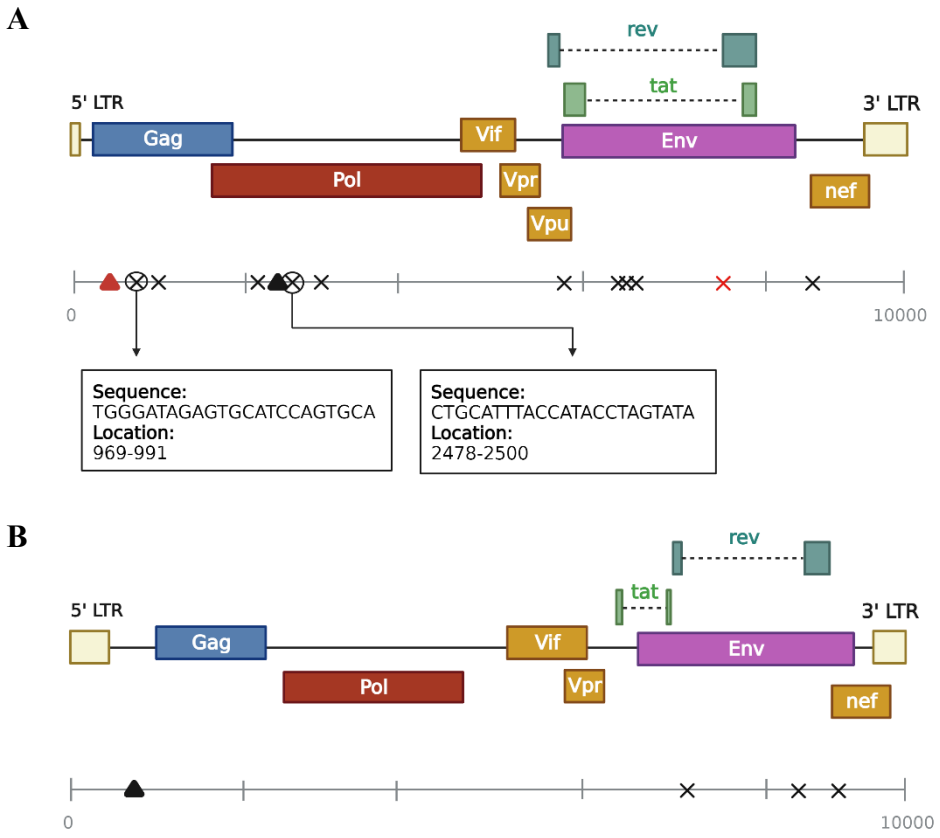


Figure 3.7. HIV-encoded miRNA-like molecules potentially regulate relevant viral genes.
(A) Schematic representation of the HIV-1 genome showing the putative miRNAs and their predicted binding sites. Triangles represent putative miRNAs (red for sncHIV1_1 and black for sncHIV1_2) and crosses symbolize their respective targets. The targets predicted at the same time and in the same position by both algorithms (Miranda and RNA22) are highlighted with a circle. (B) Schematic representation of the HIV-2 genome showing the putative miRNA sncHIV2_1 (triangle) and its targets (crosses). Created with BioRender.com

The previously used algorithms (Miranda and RNA22) were selected since, unlike most target prediction algorithms, they allow the prediction of miRNA targets in the viral genome. However, for the prediction of targets in the human genome we decided to use different tools for the following reasons: (1) The Miranda software does not have a very accessible interface, making it difficult to analyze the results in a short period. (2) RNA22 was pointed out as a low-performance algorithm compared to other more recent methods available[78–80].

We then chose to use the following online tools: Target Scan[68], an algorithm that searches for seed match and sequence conservation and presents a very high level of performance, specificity, and sensitivity, equivalent to Miranda[78–81]. miRDB[70], an algorithm that uses the same criteria as the previously described tool, together with free energy and site accessibility, thus allowing to complement the analysis with different features, and have also been shown to have high levels of precision and specificity[78,79]. Target Rank[69], which scores the seed matches in a UTR relative to a given miRNA, and then calculates an overall score for the mRNA by summing the scores for all seed matches present in the 3'UTR using seed match, conservation, target-site abundance, and local AU content features.

Although there is not much evidence comparing the performance levels of this tool with other commonly used tools, TargetRank appears to be able to predict more targets than RNA22 with equivalent levels of specificity and sensitivity[78,82].

The lists of targets predicted by at least two algorithms were used for all the analyses described in the following paragraphs. We identified 425 and 75 predicted targets for sncHIV1_1 and sncHIV1_2, respectively (Figure 3.8A/B; Table 6.1), of which 2% (10 targets) were shared (Figure 6.2). For sncHIV2_1, the target analysis retrieved a list of 134 host genes predicted by at least two of the mentioned algorithms (Figure 3.8C; Table 6.2).

We further evaluated whether our HIV-encoded miR-like sequences may act as orthologs of cellular miRNAs by looking for sequence homologies with human-derived miRNAs. None of our miRNA-like molecules shared their seed sequence with known human miRNAs listed in miRBase Release 22.1[22].

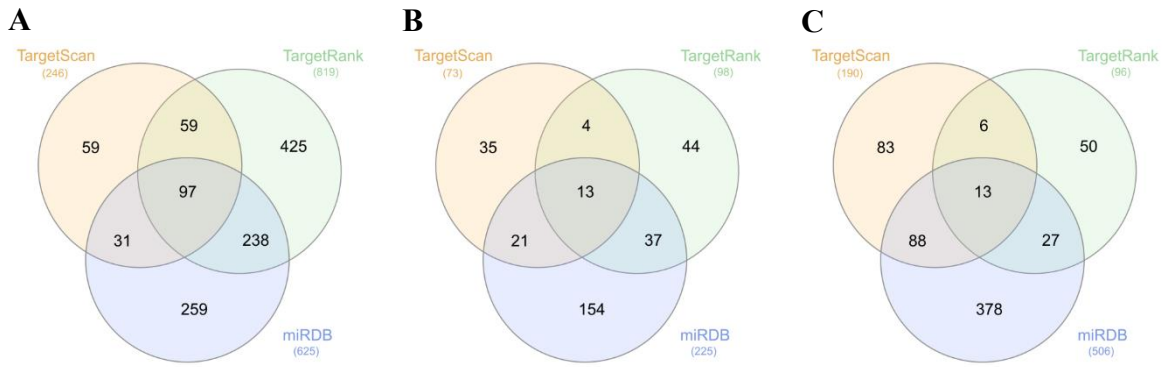


Figure 3.8. Target prediction for HIV miRNA-like molecules.
Venn diagram of genes targeted by ssncHIV1_1 (A), sncHIV1_2 (B), and sncHIV2_1 (C). Target Scan, Target Rank, and miRDB were used to predict putative miRNA targets.

3.4.2. Analysis of the interaction between predicted targets and HIV-1 proteins

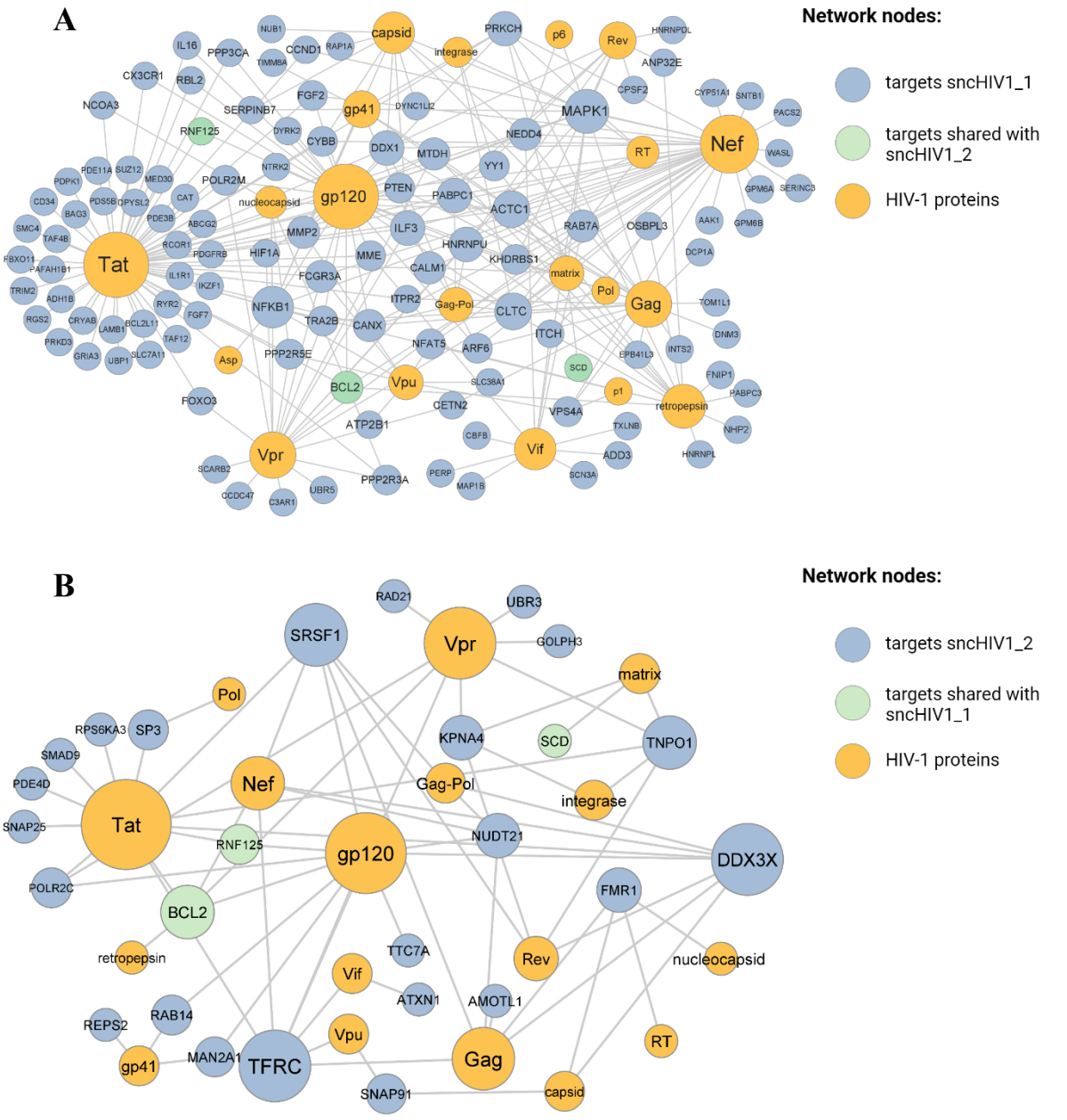
Thousands of publications have addressed the interaction of HIV-1 proteins with human host proteins. In 2008, the Division of Acquired Immunodeficiency Syndrome (DAIDS) of the National Institute of Allergy and Infectious Diseases (NIAID) compiled all of these published articles and created a searchable platform (https://www.ncbi.nlm.nih.gov/genome/viruses/retroviruses/hiv-1/interactions/) to catalog the interactions of individual HIV proteins with host cell proteins. With this platform, they were able to assess that at the date of the publication, 1434 human genes encoding 1448 proteins had been shown to directly or indirectly interact with HIV-1 proteins[72]. We analyzed this database and discovered that at present, of the total of more than 19000 genes encoding proteins existent in the human genome[83], this platform comprises 4667 human genes encoding 3859 proteins that interact with HIV-1 proteins.

Having the HIV-1 Interactome database as a starting point, we were interested in understanding whether our predicted targets displayed any direct or indirect interaction with viral proteins. For this purpose, we used the target genes predicted by at least two algorithms for sncHIV1_1 and sncHIV1_2 (shown in Table 6.1). The outcome of this analysis is represented in Figure 3.9.

Our results showed that ~27% (115 genes) and ~35% (26 genes) of our predicted targets for sncHIV1_1 and sncHIV1_2, respectively, are part of the Human-HIV-1 Interactome. Interestingly, using hypergeometric distribution analysis on R software V4.2.1 we determined that these percentages are significantly higher than expected by chance alone ($p<0,0005$ for sncHIV1_1 and $p<0,005$ for sncHIV1_2), considering the percentage of human proteins that are known to interact with HIV proteins (24%). It is also interesting to note that of the 10 targets that had been predicted in common for both candidate miRNAs, three (BCL2, RNF125, and SCD) are part of the HIV-1 interactome.

To better understand which putative targets or viral proteins are more relevant, we searched for the nodes with the highest degree of connectivity within the subnetwork that captures the predicted targets. Our data indicate that the main interactions of our predicted targets involve Tat (63 interactions for sncHIV1_1 and 12 interactions for sncHIV1_2) followed by Env gp120 (31 interactions for sncHIV1_1 and 10 interactions for sncHIV1_2). Curiously, Env gp120 was also one of the predicted binding sites in the HIV genome for both sncHIV1_1 and HIV1_2. Furthermore, MAPK1, NFKB1, CLTC, and ACTC1 for sncHIV1_1, and TRFC, DDX3X, and SRSF1 for sncHIV1_2 are the putative targets with the highest degree of connections (Figure3.9A/B).

Figure 3.9. Network of interactions between putative targets for HIV-1-encoded miRNA-like molecules and HIV-1 proteins.



Blue nodes represent specific predicted targets for sncHIV1_1 (A) and sncHIV1_2 (B), green nodes represent targets predicted for both putative miRNAs, and yellow nodes represent viral proteins. The size of the nodes is organized according to the number of interactions in the network with bigger nodes representing a higher number of interactions. The targets here represented were predicted by at least two different algorithms (TargetScan; TargetRank; miRDB).

Regarding sncHIV2_1, the Human-HIV-2 interactome has yet to be generated. Nevertheless, the high degree of homology with the HIV-1 genome (Figure 3.7A/B) justifies a similar analysis based on the Human-HIV-1 interactome. The results showed that only 32 predicted targets (~24%) belong to the HIV-1 interactome (Figure 3.10), the same percentage found for the total number of human protein-coding genes present in the interactome. Interestingly, MAPK1, similar to what was observed for sncHIV1_1, reappears as the target with the highest degree of connections (13), followed by FOS (6), and once more, Tat was the viral protein with more interactions (12), followed by Envgp120 and Nef, both with 8 connections.

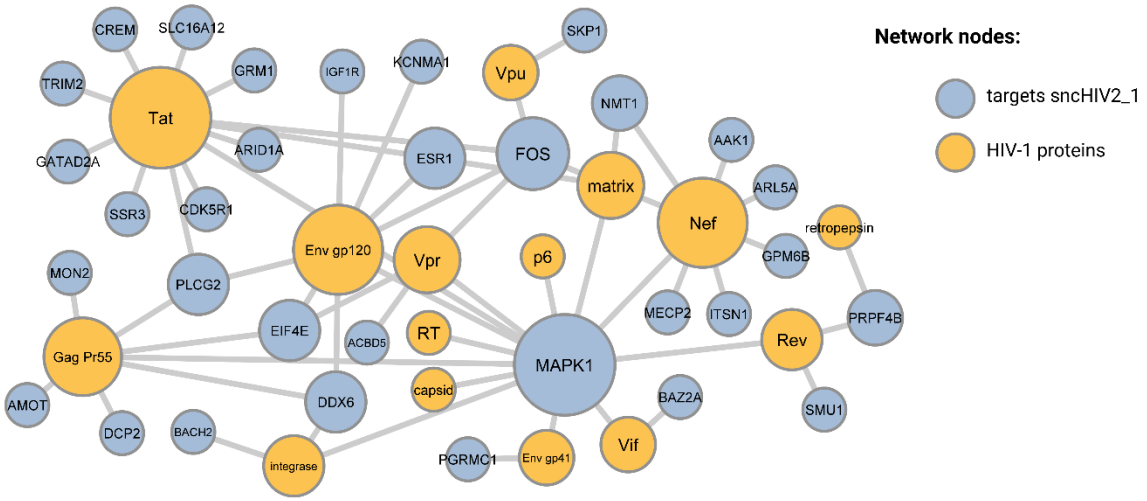


Figure 3.10. Network of interactions between putative targets for sncHIV2_1 and HIV-1 proteins.
Blue nodes represent predicted targets for sncHIV2_1 and yellow nodes represent viral proteins. The size of the nodes is organized according to the number of interactions in the network with bigger nodes representing a higher number of interactions. The targets here represented were predicted by at least two different algorithms (TargetScan; TargetRank; miRDB).

We further hypothesize that the expression of a viral miRNA would be more advantageous in contexts where it will promote the function of viral proteins. Thus, we specifically searched the interactome database for reported interactions involving the predicted targets that negatively affect the two most highly connected viral proteins (Tat and Env gp120), whether by inhibiting or downregulating them. For HIV-1-encoded miRNAs-like molecules, we found that Tat is inhibited and/or downregulated by 9 putative targets of sncHIV1_1 and 3 putative targets of sncHIV1_2 (Figure 3.11). Interestingly, BCL2 seems to be the only common target between the two miRNAs-like molecules that negatively affect TAT. Regarding Env gp120, our analysis showed that 4 putative targets for sncHIV1_1 and 1 for sncHIV1_2 downregulate or inhibit Env gp120 expression (Figure 3.11), while only NEDD4, a sncHIV1_1 target, seems to upregulate this protein (Figure 3.11).

The general proportion of negative and positive interactions in our interactome subnetwork was comparable to that found in the general HIV-1 interactome. However, for specific viral proteins, our subnetwork captured a higher proportion of human interactome partners with reported negative regulatory effects. This implies that the putative miRNAs are predominantly targeting inhibitors of viral replication, in agreement with our hypothesis.

A separate analysis was performed for sncHIV2_1. Curiously, no direct positive or negative interactions were annotated for the targets of this molecule (data not shown). The implication of this observation is limited and influenced by the fact that we are performing the analysis on the HIV-1 interactome, and there may be some differences for the HIV-2 interactome. Furthermore, we cannot ignore that an analysis of the undirected interactions described in the database (e.g., "regulated by",

"associated with", and "requires") would be helpful to better understand the role of these molecules in the respective viral genomes.

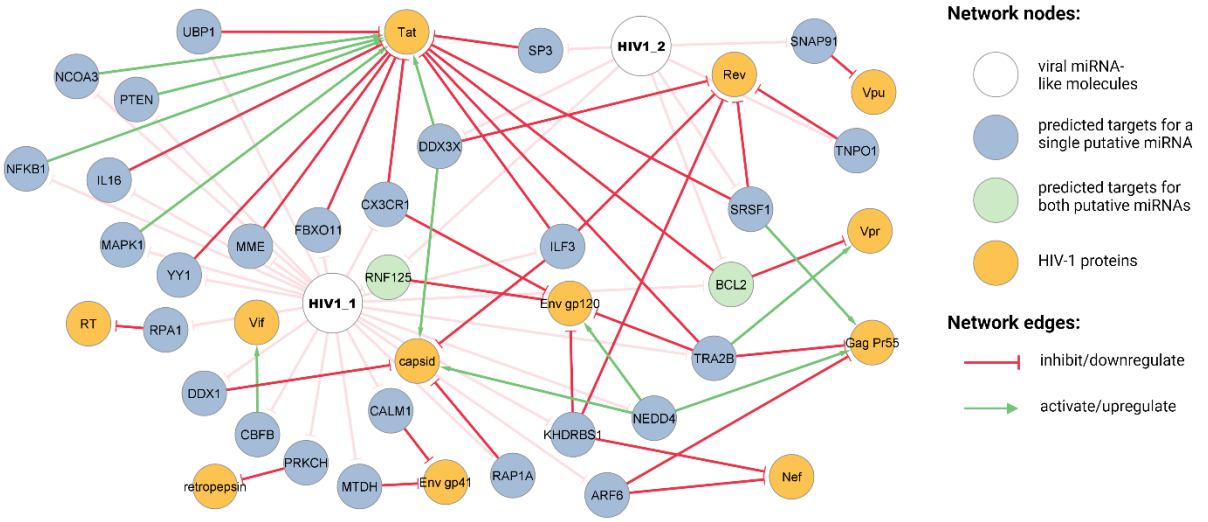


Figure 3.11. Predicted targets of HIV-1 encoded miRNA-like molecules positively and negatively regulate viral proteins. Blue nodes represent specific predicted targets for sncHIV1_1 or sncHIV1_2, green nodes represent targets predicted for both putative miRNAs, yellow nodes represent viral proteins, and with nodes represent putative miRNAs. Edges indicate the positive (green) or negative (red) effect of the predicted targets on viral proteins. The targets here represented were predicted by at least two different algorithms (TargetScan; TargetRank; miRDB).

Finally, we use the VirmiRNA database[20], which contains information about experimentally validated viral miRNAs and their targets, to search for common targets between our molecules and other viral miRNAs. We found that from the list of targets predicted by at least two algorithms, 199 and 35 targets for sncHIV1_1 and sncHIV1_2, respectively, were also described as targets of other viral miRNAs. For sncHIV2_1 we found 36 targets in common with other viral miRNAs.

For example, sncHIV1_1, sncHIV2_1, and MDV1-mir-M4-5p, a miRNA encoded by Marek's disease virus (MDVI), seem to target GPM6B, described as an oncogene and cell proliferator regulator[84,85], and similar to KSHV-encoded miRNAs[86], sncHIV1_1 may potential regulate global epigenetic reprogramming and promote viral latency by downregulation of predicted target Rbl2. Likewise, sncHIV1_2 interacts with SP3, a host gene, involved in T cell activation and immune system development, and was previously predicted as a target for HPV16-mir-H2-1[87].

In summary, this preliminary analysis revealed that there is a prevalence of predicted targets for HIV-1-encoded miRNAs-like molecules with a negative impact on viral proteins, which supports our hypothesis that the expression of our candidate viral miRNAs could promote infection through the inhibition of host proteins that negatively affect the viral proteins. Additionally, the existence of shared targets between or three candidate miRs and other viral miRNAs strengthens the idea that our molecules could play a role in the replication and maintenance of the viral cycle. To deepen the knowledge about the mechanisms of action of our putative miRNAs, the next steps, which were not possible to carry out in this work, would be to experimentally validate these predictions, for example by evaluating the expression of targets in the absence and/or over-expression of our molecules.

3.4.3. Functional enrichment of novel miRNAs-dependent transcriptome

To reveal the biological significance of the HIV-encoded sncRNAs, we were interested in understanding, on the one hand, which biological processes our targets participate in and, on the other

hand, whether there would be any GO terms and/or pathways that would be enriched in our set of targets compared to the total HIV-1 interactome (background universe).

We started by evaluating which biological processes and molecular functions our targets play a role in. This analysis was performed using PANTHER algorithm[88] and the list of total predicted targets for each candidate (targets present and not present in the interactome) was used.

The results showed that all the target genes for the three putative miRNAs are annotated and are mainly related to cellular process, biological regulation, and metabolic processes. Regarding molecular functions, 420 were identified for the sncHIV1_1 targets, 62 for the sncHIV1_2 targets, and 128 for the sncHIV2_1 targets. For biological processes, the predicted targets for sncHIV1_1, sncHIV1_2, and sncHIV2_1 showed to be associated with a total of 912, 178, and 297 processes, respectively.

Interestingly, the targets predicted for sncHIV1_1 are the only ones that show participation in processes related to the immune system sufficiently representative to appear in the analysis (Figure 3.12A). However, in the total list of biological processes in which predicted targets participate, we also find some immune-related processes associated with HIV2_1 predicted targets. The gene ontology distribution of the predicted targets for the top high-level biological processes and molecular functions are shown in Figure 3.12.

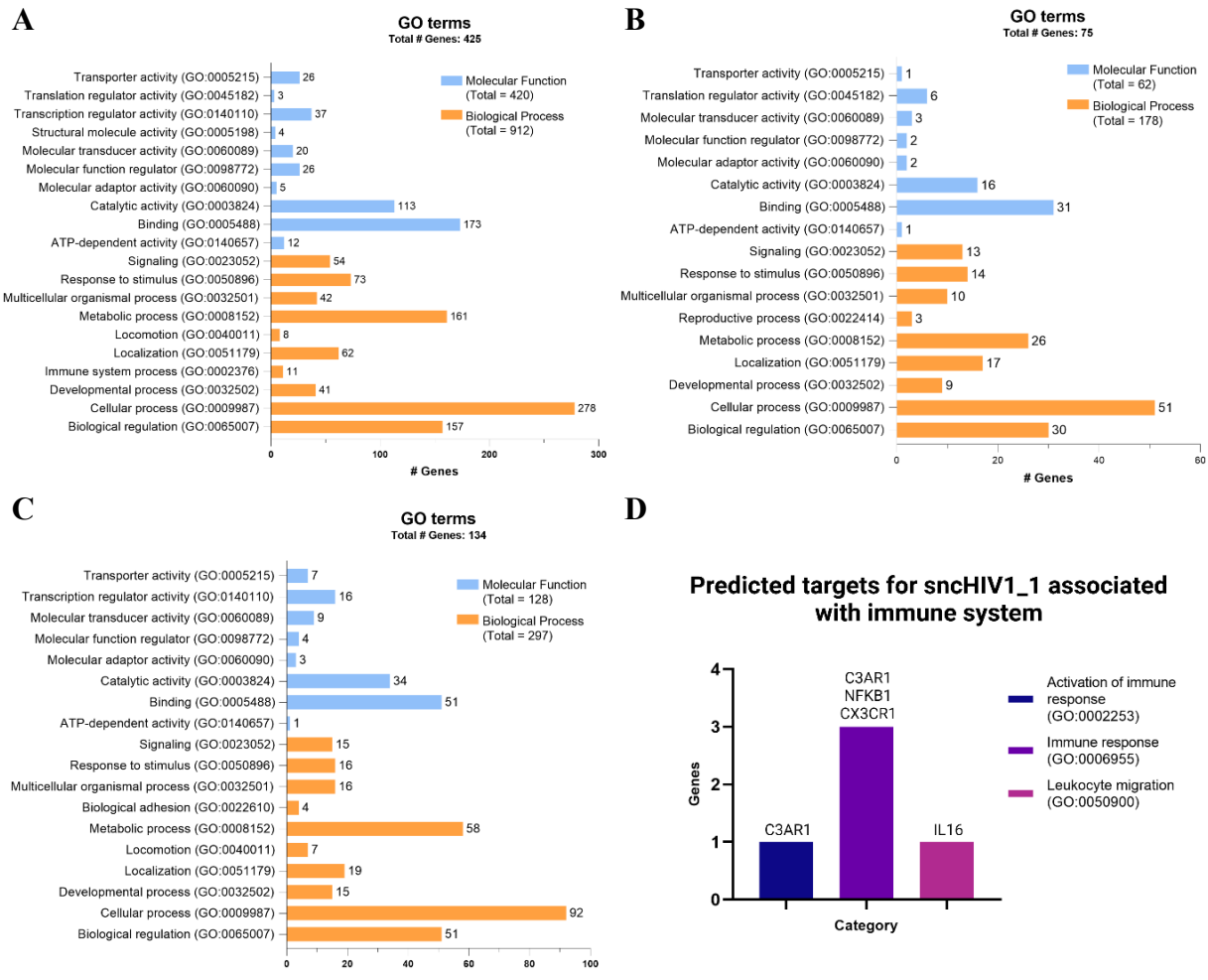


Figure 3.12. Gene ontology distribution of the predicted targets
Evaluation of the biological processes (orange bars) and molecular functions (orange bars) where predicted targets for sncHIV1_1 (A), sncHIV1_2 (B), and sncHIV2_1 (C) play a role. Only the high-level GO terms that present a more expressive representation in our set of genes are represented. (D) Predicted targets of sncHIV1_1 molecule with a potential role in immune system-related biological processes.

We repeated the analysis only for the targets present in the human interactome and the results were similar (Table 3.2). Predicted targets were mainly linked to cellular process and metabolic process, followed by biological regulation for sncHIV1_1 and sncHIV2_1 targets, or localization for sncHIV1_2 targets (Data not shown). Once again, only sncHIV1_1 targets were shown to be strongly associated with immune system functions, with 5 genes playing roles in this biological process (Figure 3.12D).

Table 3.2. Synthesis of the number of biological processes and molecular functions in which the predicted targets for our candidate miRNAs participate.

HIV-encoded miRNA-like molecule	# Predicted targets		# Predicted targets with annotations		# Molecular functions		# Biological processes	
	Total	Interactome	Total	Interactome	Total	Interactome	Total	Interactome
sncHIV1_1	425	115	425	115	420	133	912	289
sncHIV1_2	75	26	75	26	62	26	176	75
sncHIV2_1	134	32	134	32	128	39	297	88

The above analysis only gives us a general idea of the biological processes in which our predicted targets are involved. Next, we analyzed the functional enrichment of our lists of predicted targets present in the HIV-1 interactome against the universe of genes present in the HIV-1 interactome using the ShinyGO software[73] to identify statistically significant (FDR<0,05) biological processes, cellular component, molecular function, and pathways associated with our putative targets. ShinyGO was used for this analysis as an alternative to the Panther algorithm since the first one gives us the possibility to extract the results directly in graph format.

Regarding sncHIV1_1, several GO terms significantly enriched for all three categories of GO terms were found (Figure 3.13A; Table 6.3), with emphasis on the association with the response to the nerve growth factor (NGF), also strengthened by the participation of these targets in cellular components related to the nervous system, where this factor plays its main role. In fact, it is known that HIV-1 infection has the potential to cause neuronal damage, although the mechanisms of this neurotoxicity are not well understood since neurons are not directly infected by the virus[89]. Nonetheless, the protective effect of NGF against Vpr-induced toxic effects in HIV-infected patients was already demonstrated. These results may indicate that sncHIV1_1 predicted targets have a possible association with the biological mechanisms behind neurological damages during HIV-1 infection.

For sncHIV1_2 predicted targets, we were not able to find any enriched biological process or cellular component. However, molecular functions such as RNA strand annealing activity, poly(G) biding, annealing activity, RNA stem-loop binding, and poly-purine tract biding seem to be overrepresented in our gene set (Figure3.13B; Table 6.4). Interestingly, G-quadruplexes have been shown to control the expression of key viral proteins and seem to be present in short RNA templates from the HIV-1 genome, near the central poly-purine tract (cPPT) at the 3' end of the pol gene and have been implicated in the control of HIV-1 replication[90]. Moreover, a recent study showed that these G-rich regions can regulate promoter regions of HCMV-encoded miRNAs and suggest that these DNA secondary structures are a potential regulatory element of herpesvirus-encoded miRNAs[91]. The enrichment of sncHIV1_2 targets for molecular functions related to G-quadruplexes may suggest an eventual regulatory pathway of HIV infection by these target genes through modulation of G-rich regions.

Concerning sncHIV2_1, the GO classification analysis, shown in Figure 3.14 and Table 6.5, suggests that DNA alkylation and DNA methylation were the most enriched biological processes, two RNA modifications associated with the regulation of HIV-1 infection and latency[92]. In addition, the chromatoid body, RISC complex, and RNAi effector complex were the cellular compartment with higher fold enrichment, followed by post synapse and chromatin. Of note, although the molecular function of the chromatoid body is still very debated, there is evidence that these structures, present in the male germ line, act as an RNA storing and processing region and accumulate key proteins of the

miRNA machinery such as Dicer and AGO[93]. Interestingly, when we repeated the analysis for the full set of predicted targets (and not just those present in Interactome) the results were slightly different.

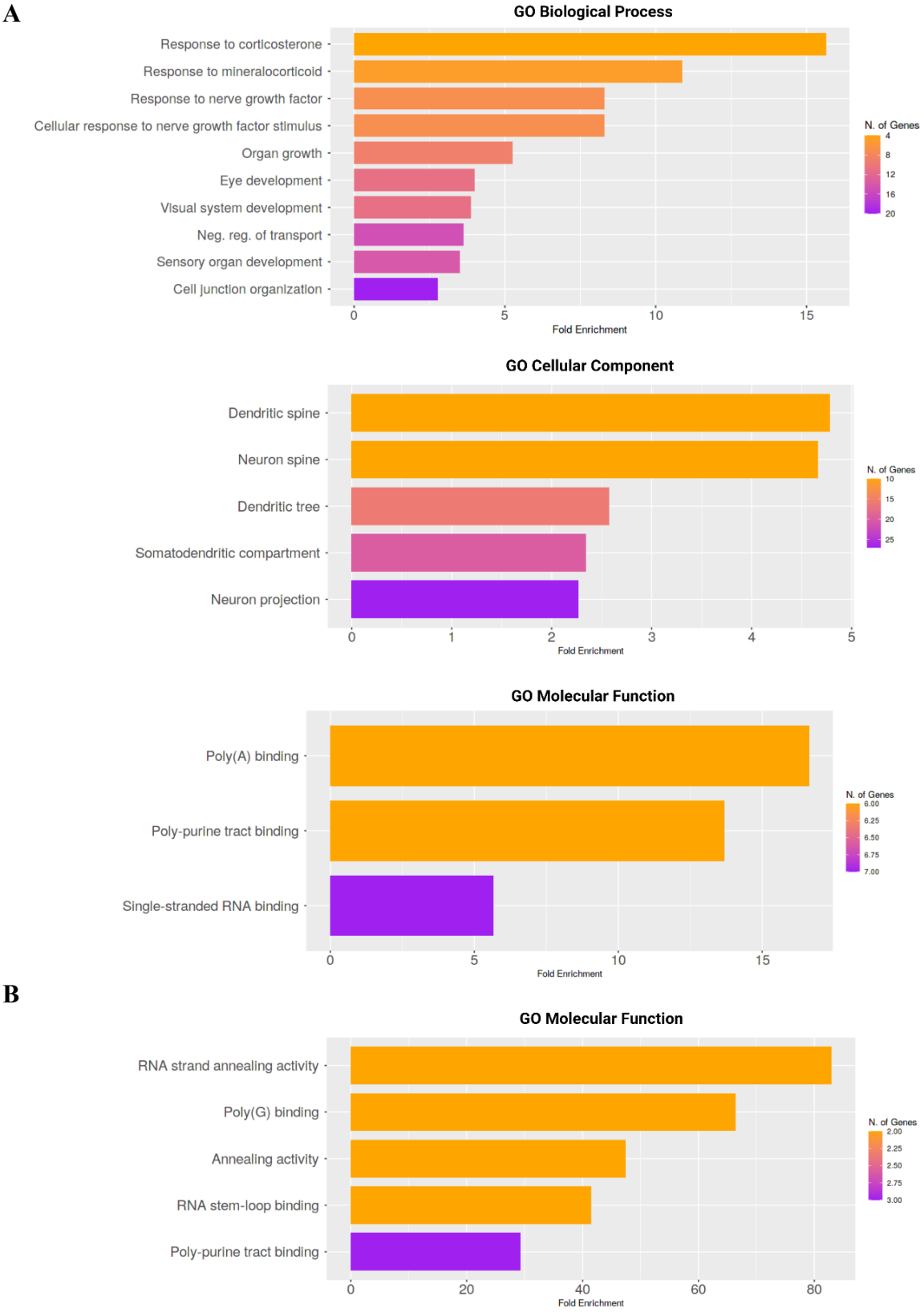


Figure 3.13. Functional enrichment analysis of genes predicted as targets for HIV-1-encoded miRNA-like molecules
(A) A gene ontology (GO) enrichment analysis of the targets predicted for sncHIV1_1 show several enriched biological processes, cellular components, and molecular functions. (B) sncHIV1_2 targets are not enriched for any biological process or cellular component but show enrichment for various molecular functions.

For the predicted target list of sncHIV1_1, we observed a decrease in the number of enriched biological processes, with the majority being related to the development of the ocular system, no enriched cellular components, and an increase in the number of enriched molecular functions (data not shown). For the predicted targets of the two remaining molecules, when extending the analysis to targets not present in the interactome, no enrichments were found for the three categories of GO terms.

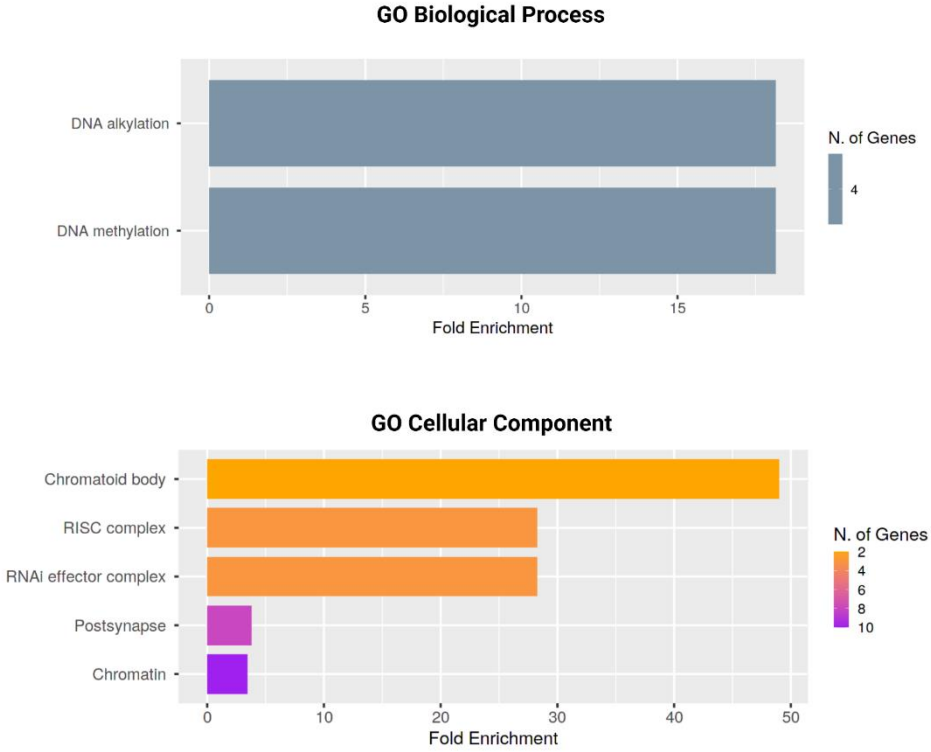


Figure 3.14. Functional enrichment analysis of genes predicted as targets for sncHIV2_1 molecule.
A Gene ontology (GO) enrichment analysis of the targets predicted for sncHIV2_1 show a few enriched biological processes, and cellular components but no molecular function.

To further explore the functional impact of our molecules, we searched for possibly enriched pathways for our sncRNAs targets. Since no enriched KEGG pathways were found for any of our three putative miRNAs predicted targets, we decide to alternatively use the Panther curated pathways.

For sncHIV2_1 targets we did not find any significantly enriched pathway, however, the analysis retrieved one enriched pathway for both sncHIV1_1 and sncHIV2_1 predicted targets. The putative miR HIV1_1 predicted targets seem to be associated with the CCKR signaling map, a poorly understood pathway that has been linked with cell survival and proliferation with a possible carcinogenic role[94,95]. Moreover, Insulin/IGF pathway-mitogen activated protein kinase B signaling cascade is overrepresented for sncHIV2_1 targets. This pathway is vastly associated with a variety of cellular processes, including cell proliferation and apoptosis[96]. In summary, both molecules appear to have targets that play a role in cell survival and proliferation, two canonical mechanisms associated with viral-encoded miRNAs to enhance viral replication.

Analysis of the panther pathways enriched for the set of genes present and not present in the HIV-1 interactome and predicted as targets for each of our candidates did not retrieve any results.

Overall, our data support the possibility that both HIV-1 and HIV-2 are able to encode miRNA-like molecules, produced through Dicer-dependent processing, with the potential to modulate pathways critical for viral replication and immune responses.

4. DISCUSSION

miRNAs are small non-coding RNAs crucial in the post-transcriptional regulation of host and viral gene expression. Besides host miRNAs, which have been extensively studied and associated with a broad range of functions such as cell proliferation, cell differentiation, and anti-viral protection, the ability of viruses to encoded miRNAs has also been demonstrated, with most v-miRNAs being encoded by DNA viruses.

On other hand, although some miRNAs produced by RNA viruses have already been validated, until very recently it was assumed that these viruses would not be able to encode miRNAs. Accordingly, the existence of HIV-1 encoded miRNAs has been controversial since the identification of the first HIV-encoded miR-like molecule in 2004. It has been argued that the production of a canonical pre-miRNA would result in the wasteful cleavage and, ultimately, destruction of viral genomic RNAs, thus reducing viral replication. However, recent studies have shown that different retroviruses, including HIV-1, can encode miRNAs using both canonical and non-canonical pathways. Even so, the miRNAs identified for HIV-1 were mostly predicted through bioinformatic analysis and some do not have the typical hairpin structure of true miRNAs, which makes the validation of these molecules controversial. The ability of HIV-2 to produce miRNAs has never been addressed.

Previous work from our lab used small-RNA-seq libraries derived from *in vitro* TCR-stimulated naïve CD4+ T cells isolated from healthy blood donors and infected with either HIV-1 or HIV-2 clones to search for HIV-specific reads corresponding to sncRNAs. This previous work identified three strong miRNAs candidates, two corresponding to HIV-1, here named sncHIV1_1 and sncHIV1_2, and one corresponding to HIV-1 (sncHIV2_1) with a high level of conservation and a stem-loop-like secondary structure. However, experimental validation would be necessary to demonstrate that these molecules are true, functional miRNAs, rather than "noise" from degradation products.

In this work, we integrated distinct approaches focusing on human cell models and bioinformatic tools to experimentally validate these candidate miRNAs and understand their role in viral infection and replication. In order to experimentally validate putative miRNAs, our first question was to determine if the mature miRNA sequence would accumulate in cells transfected with the plasmid vector encoding the pre-miR. For this purpose, we took advantage of previously cloned pre-miRNAs available in the lab and performed an RT-qPCR stem-loop method to evaluate the expression of our molecules. The results showed that sncHIV1_2 and sncHIV2_1 are successfully expressed in HEK cells transfected with a plasmid encoding the predicted pre-miR sequence (Figure 3.2A/B). However, we were unable to detect sncHIV1_1, for which we could only amplify low levels of non-specific products. Despite this, we decided to continue with the analysis of the three sequences. Given the nature of the RT-qPCR method used, we cannot exclude at the moment that our inability to successfully amplify sncHIV1_1 results from a problem with the primer sequences.

Knowing that miRNA detection methods based on RT-qPCR do not enable the simultaneous detection of primary, precursor, and mature forms of miRNAs, nor the direct visualization of the transcript size, we tried to implement in the laboratory a detection method based on northern blot hybridization[65]. However, despite the preliminary results of the establishment of this technique, we faced specificity problems that we were unable to resolve in the time frame of this work.

The next question to consider was whether the formation of a hairpin structure is required for the expression of the putative miRNA. To address this, we took advantage of site-directed mutagenesis on the wild-type sequences to introduce mutations predicted to disrupt the hairpin structure and analyzed their impact on the levels of mature miR by RT-qPCR. The quantification revealed that mutant sequences lead to a significant decrease in the accumulation of mature miRs sncHIV1_2 and

sncHIV2_1. As before, the candidate miR HIV1_1 was only detected at very low levels and no significant difference was observed between mutant and wildtype vectors (Figure 3.4).

To assess whether HIV-encoded miRs were a product of the Dicer processing, we further used a HEK 293T cell line inducible for shRNA knock-down of DICER. Since we could not establish a method in time to validate the knock-down of Dicer in these cells at the protein level, we assume that the results obtained for the mRNA levels reflect what is happening at the protein level. Our results showed that Dicer knockdown led to a significant decline in sncHIV1_2 and sncHIV2_1 mature levels (Figure 3.6). Of note, the decrease in relative viral miRNA levels of these putative miRNAs in cells with the Dicer knock-down was similar to that observed for mir34c-5p, a *bonafide* miRNA used as a control.

Together these results provide forceful evidence to support at least two of our molecules (sncHIV1_2 and sncHIV2_1) as strong *bonafide* miRNA candidates produced by Dicer-dependent processing, rather than degradation products. Additional investigations will be needed to determine the mechanisms by which HIV-1 and HIV-2 encode miRNAs. In addition to the dependence that our candidates have on DICER, validated in this work, it would be interesting to determine whether other key steps in the canonical biosynthesis of miRNAs, such as processing by DROSHA, transport by Exportin-5, or accumulation in AGO, participate in the process of synthesis of our molecules. Since previous studies failed to detect the incorporation of HIV-1-derived sncRNAs into the AGO-RISC complex[60], we believe that this is one of the key miRNA biosynthesis steps to be tested in future work. For this, we have already devised a strategy based on tandem affinity purification. This approach described by Nonne et al.[97] consists of two steps. Briefly, cells containing a tagged version of Argonaute are immunoprecipitated using anti-FLAG antibodies and then the miRNAs associated with the protein are affinity purified using streptavidin beads.

To further dissect the functional impact of candidate miRNAs during infection and host-virus interaction, we performed a preliminary prediction of the HIV candidate-miR target genes in host and virus genomes. These analyses provided some clues for a better understanding of these molecules' possible functional roles. Using stringent evaluation criteria and different algorithms of target prediction we demonstrated that sncHIV1_1, sncHIV1_2, and sncHIV2_1, were all potentially able to target a variety of both viral and cellular targets. Moreover, we showed that a high percentage of the predicted targets are present in the HIV-1 interactome. We found a prevalence of predicted targets for HIV-1-encoded miRNAs-like molecules with a negative impact on viral proteins, which supports our hypothesis that the expression of our candidate viral miRNAs could promote infection through the inhibition of host proteins that negatively affect the viral proteins. Analysis of the connectivity of the predicted targets within the HIV-1 interactome (Figure 3.9 and Figure 3.10) allowed us to determine which ones have the highest impact on host-virus interactions (major nodes) when potentially downregulated by HIV miRNAs. We identified MAPK1 and NFKB1, CLTC, and ACTC1 as major nodes for sncHIV1_1. MAPK1 participates in the MAPK pathway, highly associated with cell proliferation and apoptosis, and has been shown to be downregulated in HCV-infected cells leading to an increase in HCV replication[98]. NFKB1 is a key component of the NF-κB canonical pathway. The NFKB pathway regulates survival, differentiation, and activation of T cell activation[99], which in turn is a requirement for viral replication. CLTC is involved in clathrin-mediated endocytosis that seems to regulate HIV-1 transfer to T cells[100]. ACTC1 is a member of actin cytoskeletons, reported to potentially regulate lung tumors, and found significantly downregulated in virus-infected cells[101]. Regarding sncHIV1_2, we identified TFRC, DDX3X, and SRSF1 as major nodes. DDX3X and SRF1 both downregulate REV activity[102,103]. Moreover, SRF1 also impairs TAT expression and has been shown to activate transcription in the early stages of viral infection and be downregulated in later stages[104].TFRC encodes cell surface receptors required for neurologic development and participates in iron uptake processes[105]. Thus, our data suggest that the HIV-1 encoded miRNAs we identified could act synergistically, targeting mRNAs that are involved in controlling HIV replication. However, only one

of these molecules (sncHIV1_2) was validated by the experimental methods used in this work. The interesting targets found for sncHIV1_1, for which it was not possible to obtain experimental validation, indicate that the difficulties of experimental validation may be due to technical problems of the PCR method that needs optimization. Since these targets are also required for other functions of the virus, being activated by different viral proteins, the expression of such viral miRNAs requires a fine-tuning regulation as is the case for several reported v- miRNAs (recently reviewed in depth [19]), an area that warrants further investigation.

Regarding the potential miRNA-like molecule identified in HIV-2, we found that it shared MAPK1 as a major target with snchHIV1_1, followed by FOS. FOS proteins have been associated with cell proliferation, transformation, and apoptosis and present several interactions within the viral interactome. Moreover, a study reported that the HCMV-encoded viral G protein-coupled receptor (pUS28) promotes viral latency by attenuation of c-fos expression in HCMV-infected cells[106]. Given our results, we may therefore hypothesize that HIV-2 putative miRNAs may potentially regulate host genes through a miRNA-mediated pathway, modulating the immune response as seen for HIV-1 encoded miR-like molecules.

We further showed that sncHIV1_1, sncHIV1_2, and sncHIV2_1 predicted targets displayed a significant enrichment for several biological processes, molecular functions, cellular components, and pathways in comparison to the universe of genes present in the HIV-1 interactome. The list of predicted targets for sncHIV1_1 showed significant enrichment of genes involved in neuronal processes with emphasis on the association with NGF and corticosterone, strengthened by the participation of these targets in cellular components related to the nervous system such as dendritic spine, neuron spine, dendritic tree, somatodendritic compartment, and neuron projection (Figure 13.3A and Table 6.3). The NGF plays an essential role in the survival, differentiation, and regeneration of neurons in the peripheral nervous system[89]. The protective effect of NGF against Vpr-induced toxic effects in HIV-infected patients was already demonstrated thus making NGF a potential therapeutic target for the mitigation of HIV-1-directed toxicity in the brain[89]. Corticosterone also seems to have a role in neurological-related processes. A study showed that transgenic mice models that mimic the HIV-1 envelope glycoprotein gp120-induced nerve damage present an elevated level of corticosteron[107]. This indicates that HIV-encoded sncHIV1_1 predicted targets may play a role in HIV-1-induced neurotoxicity for which the underlying mechanisms are not well understood. Additionally, sncHIV1_1 predicted targets seem to be associated with the CCKR signaling map, a poorly understood pathway that has been linked with cell survival and proliferation with a possible carcinogenic role[94,95].

sncHIV1_2 predicted targets showed significant enrichment for molecular functions related to RNA strand annealing activity, poly(G) biding, RNA stem-loop binding, annealing activity, and poly-purine tract biding seem to be overrepresented in our gene set (Figure3.13B; Table 6.4). G-quadruplexes have been shown to control the expression of key viral proteins and control HIV-1 replication[90]. Moreover, a recent study showed that these G-rich regions can regulate promoter regions of HCMV-encoded miRNAs and suggest that these DNA secondary structures are a potential regulatory element of herpesvirus-encoded miRNAs[91]. Thus, our results suggest that sncHIV1_2 predicted targets may be implicated in an eventual regulatory pathway of HIV infection through modulation of G-rich regions and other important binding processes. Moreover, Insulin/IGF pathway-mitogen activated protein kinase B signaling cascade, a pathway that regulates cell proliferation and apoptosis[96], is overrepresented for sncHIV2_1 predicted targets.

Concerning sncHIV2_1 targets, the GO classification analysis showed significant enrichment in RNA modification processes (alkylation and methylation), normally associated with the regulation of HIV-1 infection and latency[92].Several cellular compartments associated with RNA metabolism and synthesis (chromatoid body, RISC complex, and RNAi effector complex) also showed a high fold enrichment, with a particular interest for chromatoid body, structures present in the male germ line that act as an

RNA storing and processing region and accumulate key proteins of the miRNA machinery such as Dicer and AGO[93].

Taken together, these results open the door to understanding the mechanism of action of our candidate miRNAs during viral infection. The predicted targets for the three molecules show interesting connections to important biological processes that regulate viral infection, latency, and cell survival. Importantly, the targets of the three molecules do not seem to be related to the same mechanisms, which demonstrates that HIV may have the ability to synthesize miRNAs that differentially regulate different pathways to promote viral infection and replication. However, it is important to note that both putative HIV-1 miRNAs appear to be associated with biological mechanisms that regulate proliferation, cell survival, and apoptosis as observed for other cellular and viral miRNAs (reviewed in [8,19]).

In summary, the results presented in this thesis provide support for the ability of both HIV-1 and HIV-2 genomes to encode miRNA-like molecules that display the typical features of canonical miRNAs and appear to regulate host and viral gene expression. At least two of our candidate miRNAs (sncHIV1_2 and sncHIV2_1) do not appear to be RNA degradation products, but rather *bonafide* miRNAs with a hairpin structure essential for their expression and processed through a DICER-dependent mechanism.

Due to the clinical relevance that v-miRNAs have shown in recent times and the constant difficulty in controlling HIV-1 latency despite knowledge of its infection mechanisms and advances in controlling its replication, we believe that it is essential to identify and validate new miRNAs produced by this virus. For the same reason, we believe that this work constitutes an important contribution to the understanding of the mechanisms of viral infection associated with miRNAs and opens the door to future studies that intend to explore the use of these molecules for therapeutic and research purposes.

5. REFERENCES

1. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Front Endocrinol (Lausanne)* **9**, 402 (2018).

2. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to lin-14 Ambros and Horvitz, 1987). Animals carrying a lin-4 loss-of-function (lf) mutation, lin-4(e912), display reiterations of early fates at inappropriately late developmental stages; cell lineage patterns normally specific for the L1 are reiter. *Cell* **75**, 843–854 (1993).

3. Li, X. & Zou, X. An overview of rna virus-encoded micrornas. *ExRNA* **1**, 1–5 (2019).

4. L, G., R, Z., J, X., CI, W. & X, L. Functional Conservation of Both CDS- and 3'-UTR-Located MicroRNA Binding Sites between Species. *Mol Biol Evol* **32**, 3276 (2015).

5. Biswas, S. *et al.* Comparison of miRNA expression profiles between HIV-1 and HIV-2 infected monocyte-derived macrophages (MDMs) and peripheral blood mononuclear cells (PBMCs). *Int J Mol Sci* **21**, 1–25 (2020).

6. Balasubramaniam, M., Pandhare, J. & Dash, C. Are microRNAs Important Players in HIV-1 Infection? An Update. *Viruses 2018, Vol. 10, Page 110* **10**, 110 (2018).

7. Bhaskaran, M. & Mohan, M. MicroRNAs: History, Biogenesis, and Their Evolving Role in Animal Development and Disease. *Vet Pathol* **51**, 759 (2014).

8. Nanbo, A., Furuyama, W. & Lin, Z. RNA Virus-Encoded miRNAs: Current Insights and Future Challenges. *Front Microbiol* **12**, 1539 (2021).

9. Chiang, K., Liu, H. & Rice, A. P. miR-132 enhances HIV-1 replication. *Virology* **438**, 1 (2013).

10. Amaral, A. J. *et al.* miRNA profiling of human naive CD4 T cells links miR-34c-5p to cell activation and HIV replication. *EMBO J* **36**, 346–360 (2017).

11. Farberov, L. *et al.* MicroRNA-mediated regulation of p21 and TASK1 cellular restriction factors enhances HIV-1 infection. *J Cell Sci* **128**, 1607–1616 (2015).

12. Ahluwalia, J. K. *et al.* Human cellular microRNA hsa-miR-29a interferes with viral nef protein expression and HIV-1 replication. *Retrovirology* **5**, 117 (2008).

13. Nathans, R. *et al.* Cellular microRNA and P bodies modulate host-HIV-1 interactions. *Mol Cell* **34**, 696–709 (2009).

14. Sun, G. *et al.* Interplay between HIV-1 infection and host microRNAs. *Nucleic Acids Res* **40**, 2181–2196 (2012).

15. Houzet, L. *et al.* The extent of sequence complementarity correlates with the potency of cellular miRNA-mediated restriction of HIV-1. *Nucleic Acids Res* **40**, 11684–11696 (2012).

16. Joshi, N., Chandane Tak, M. & Mukherjee, A. The involvement of microRNAs in HCV and HIV infection. *Ther Adv Vaccines Immunother* **10**, (2022).

17. Klase, Z., Houzet, L. & Jeang, K. T. MicroRNAs and HIV-1: complex interactions. *J Biol Chem* **287**, 40884–40890 (2012).

18. Chinniah, R., Adimulam, T., Nandlal, L., Arumugam, T. & Ramsuran, V. The Effect of miRNA Gene Regulation on HIV Disease. *Front Genet* **13**, (2022).

19. Mishra, R., Kumar, A., Ingle, H. & Kumar, H. The Interplay Between Viral-Derived miRNAs and Host Immunity During Infection. *Front Immunol* **10**, 3079 (2020).

20. Qureshi, A., Thakur, N., Monga, I., Thakur, A. & Kumar, M. VIRmiRNA: a comprehensive resource for experimentally validated viral miRNAs and their targets. *Database (Oxford)* **2014**, (2014).

21. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140–D144 (2006).

22. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* **47**, D155–D162 (2019).

23. Morales, L. *et al.* SARS-CoV-Encoded Small RNAs Contribute to Infection-Associated Lung Pathology. *Cell Host Microbe* **21**, 344–355 (2017).

24. Aydemir, M. N. *et al.* Computationally predicted SARS-COV-2 encoded microRNAs target NFKB, JAK/STAT and TGFB signaling pathways. *Gene Rep* **22**, 101012 (2021).

25. Grundhoff, A. & Sullivan, C. S. Virus-encoded microRNAs. *Virology* **411**, 325 (2011).

26. Takane, K. & Kanai, A. Vertebrate virus-encoded microRNAs and their sequence conservation. *Jpn J Infect Dis* **64**, 357–366 (2011).

27. Boss, I. W. & Renne, R. Viral miRNAs: Tools for immune evasion. *Curr Opin Microbiol* **13**, 540–545 (2010).

28. Umbach, J. L. *et al.* MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. *Nature 2008 454:7205* **454**, 780–783 (2008).

29. Tang, S., Patel, A. & Krause, P. R. Novel Less-Abundant Viral MicroRNAs Encoded by Herpes Simplex Virus 2 Latency-Associated Transcript and Their Roles in Regulating ICP34.5 and ICP0 mRNAs. *J Virol* **83**, 1433–1442 (2009).

30. Grey, F., Meyers, H., White, E. A., Spector, D. H. & Nelson, J. A Human Cytomegalovirus-Encoded microRNA Regulates Expression of Multiple Viral Genes Involved in Replication. *PLoS Pathog* **3**, e163 (2007).

31. Lei, X. *et al.* Regulation of NF-kappaB inhibitor IkappaBalpha and viral replication by a KSHV microRNA. *Nat Cell Biol* **12**, 193–199 (2010).

32. Gottwein, E. & Cullen, B. R. A Human Herpesvirus MicroRNA Inhibits p21 Expression and Attenuates p21-Mediated Cell Cycle Arrest. *J Virol* **84**, 5229–5237 (2010).

33. Abend, J. R. *et al.* Kaposi's Sarcoma-Associated Herpesvirus MicroRNAs Target IRAK1 and MYD88, Two Components of the Toll-Like Receptor/Interleukin-1R Signaling Cascade, To Reduce Inflammatory-Cytokine Expression. *J Virol* **86**, 11663–11674 (2012).

34. Bekerman, E., Jeon, D., Ardolino, M. & Coscoy, L. A Role for Host Activation-Induced Cytidine Deaminase in Innate Immune Defense against KSHV. *PLoS Pathog* **9**, e1003748 (2013).

35. Hook, L. M. *et al.* Cytomegalovirus miRNAs Target Secretory Pathway Genes to Facilitate Formation of the Virion Assembly Compartment and Reduce Cytokine Secretion. *Cell Host Microbe* **15**, 363–373 (2014).

36. Bennasser, Y., Le, S. Y., Yeung, M. L. & Jeang, K. T. HIV-1 encoded candidate micro-RNAs and their cellular targets. *Retrovirology* **1**, (2004).

37. Klase, Z. *et al.* HIV-1 TAR miRNA protects against apoptosis by altering cellular gene expression. *Retrovirology* **6**, (2009).

38. van der Ree, M. H. *et al.* Safety, tolerability, and antiviral effect of RG-101 in patients with chronic hepatitis C: a phase 1B, double-blind, randomised controlled trial. *Lancet* **389**, 709–717 (2017).

39. Janssen, H. L. A. *et al.* Treatment of HCV infection by targeting microRNA. *N Engl J Med* **368**, 1685–1694 (2013).

40. Lu, T. *et al.* Circulating Epstein-Barr virus microRNAs BART7-3p and BART13-3p as novel biomarkers in nasopharyngeal carcinoma. *Cancer Sci* **111**, 1711–1723 (2020).

41. Pan, Y. *et al.* Circulating human cytomegalovirus-encoded HCMV-miR-US4-1 as an indicator for predicting the efficacy of IFNα treatment in chronic hepatitis B patients. *Sci Rep* **6**, (2016).

42. Fu, Z. *et al.* A virus-derived microRNA-like small RNA serves as a serum biomarker to prioritize the COVID-19 patients at high risk of developing severe disease. *Cell Discov* **7**, (2021).

43. Luciw, P. A. & Leung, N. J. Mechanisms of Retrovirus Replication. *The Retroviridae* 159–298 (1992) doi:10.1007/978-1-4615-3372-6_5.

44. Gillet, N. *et al.* Mechanisms of leukemogenesis induced by bovine leukemia virus: Prospects for novel anti-retroviral therapies in human. *Retrovirology* **4**, 1–32 (2007).

45. Kincaid, R. P., Burke, J. M. & Sullivan, C. S. RNA virus microRNA that mimics a B-cell oncomiR. *Proc Natl Acad Sci U S A* **109**, 3077–3082 (2012).

46. Whisnant, A. W. *et al.* Identification of novel, highly expressed retroviral microRNAs in cells infected by bovine foamy virus. *J Virol* **88**, 4679–4686 (2014).

47. Yao, Y., Smith, L. P., Nair, V. & Watson, M. An avian retrovirus uses canonical expression and processing mechanisms to generate viral microRNA. *J Virol* **88**, 2–9 (2014).

48. Omoto, S. *et al.* HIV-1 nef suppression by virally encoded microRNA. *Retrovirology* **1**, (2004).

49. Omoto, S. & Fujii, Y. R. Regulation of human immunodeficiency virus 1 transcription by nef microRNA. *J Gen Virol* **86**, 751–755 (2005).

50. Pfeffer, S. *et al.* Identification of microRNAs of the herpesvirus family. *Nat Methods* **2**, 269–276 (2005).

51. Lin, J. & Cullen, B. R. Analysis of the interaction of primate retroviruses with the human RNA interference machinery. *J Virol* **81**, 12218–12226 (2007).

52. Lamers Susanna L., S. L., Fogel Gary B., G. B. & McGrath, M. S. HIV-miR-H1 evolvability during HIV pathogenesis. *Biosystems* **101**, 88–96 (2010).

53. Kaul, D., Ahlawat, A. & Gupta, S. D. HIV-1 genome-encoded hiv1-mir-H1 impairs cellular responses to infection. *Mol Cell Biochem* **323**, 143–148 (2009).

54. Klase, Z. *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 Mol Biol* **8**, (2007).

55. Ouellet, D. L. *et al.* Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. *Nucleic Acids Res* **36**, 2353–2365 (2008).

56. Harwig, A., Jongejan, A., van Kampen, A. H. C., Berkhout, B. & Das, A. T. Tat-dependent production of an HIV-1 TAR-encoded miRNA-like small RNA. *Nucleic Acids Res* **44**, 4340–4353 (2016).

57. Zhang, Y. *et al.* A novel HIV-1-encoded microRNA enhances its viral replication by targeting the TATA box region. *Retrovirology* **11**, (2014).

58. Schopman, N. C. T. *et al.* Deep sequencing of virus-infected cells reveals HIV-encoded small RNAs. *Nucleic Acids Res* **40**, 414–427 (2012).

59. Whisnant, A. W. *et al.* In-depth analysis of the interaction of HIV-1 with cellular microRNA biogenesis and effector mechanisms. *mBio* **4**, (2013).

60. Vongrad, V. *et al.* HIV-1 RNAs are Not Part of the Argonaute 2 Associated RNA Interference Pathway in Macrophages. *PLoS One* **10**, (2015).

61. Simões, A. E. S. *et al.* Efficient recovery of proteins from multiple source samples after trizol® or trizol®LS RNA extraction and long-term storage. *BMC Genomics* **14**, 1–15 (2013).

62. Kramer, M. F. Stem-loop RT-qPCR for miRNAs. *Curr Protoc Mol Biol* **Chapter 15**, Unit 15.10-Unit 15.10 (2011).

63. Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E. F. & Hellens, R. P. Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods* **3**, (2007).

64. Chen, C. *et al.* Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res* **33**, (2005).

65. Ahmad, W., Gull, B., Baby, J. & Mustafa, F. A Comprehensive Analysis of Northern versus Liquid Hybridization Assays for mRNAs, Small RNAs, and miRNAs Using a Non-Radiolabeled Approach. *Curr Issues Mol Biol* **43**, 457–484 (2021).

66. John, B. *et al.* Human MicroRNA targets. *PLoS Biol* **2**, (2004).

67. Miranda, K. C. *et al.* A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. *Cell* **126**, 1203–1217 (2006).

68. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).

69. Nielsen, C. B. *et al.* Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* **13**, 1894–1910 (2007).

70. Wong, N. & Wang, X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res* **43**, D146–D152 (2015).

71. Liu, W. & Wang, X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol* **20**, (2019).

72. Fu, W. *et al.* Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res* **37**, (2009).

73. Ge, S. X., Jung, D., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628–2629 (2020).

74. Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H. & Tomari, Y. Recognition of the pre-miRNA structure by Drosophila Dicer-1. *Nat Struct Mol Biol* **18**, 1153–1158 (2011).

75. Shi, J. *et al.* Identification and validation of a novel microRNA-like molecule derived from a cytoplasmic RNA virus antigenome by bioinformatics and experimental approaches. *Virol J* **11**, 1–14 (2014).

76. Walz, N., Christalla, T., Tessmer, U. & Grundhoff, A. A global analysis of evolutionary conservation among known and predicted gammaherpesvirus microRNAs. *J Virol* **84**, 716–728 (2010).

77. Enright, A. J. *et al.* MicroRNA targets in Drosophila. *Genome Biol* **5**, 1–14 (2003).

78. Kern, F. *et al.* Validation of human microRNA target pathways enables evaluation of target prediction tools. *Nucleic Acids Res* **49**, 127–144 (2021).

79. Sturm, M., Hackenberg, M., Langenberger, D. & Frishman, D. TargetSpy: A supervised machine learning approach for microRNA target prediction. *BMC Bioinformatics* **11**, 1–17 (2010).

80. Oliveira, A. C. *et al.* Combining Results from Distinct MicroRNA Target Prediction Tools Enhances the Performance of Analyses. *Front Genet* **8**, (2017).

81. Zhang, Y. & Verbeek, F. J. Comparison and integration of target prediction algorithms for microRNA studies. *J Integr Bioinform* **7**, (2010).

82. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, (2015).

83. Piovesan, A. *et al.* Human protein-coding genes and gene feature statistics in 2019. *BMC Res Notes* **12**, 1–5 (2019).

84. Zhuang, G., Sun, A., Teng, M. & Luo, J. A tiny RNA that packs a big punch: The critical role of a viral miR-155 ortholog in Lymphomagenesis in Marek's disease. *Front Microbiol* **8**, 1169 (2017).

85. He, S. *et al.* GPM6B Inhibit PCa Proliferation by Blocking Prostate Cancer Cell Serotonin Absorptive Capacity. *Dis Markers* **2020**, (2020).

86. Lu, F., Stedman, W., Yousef, M., Renne, R. & Lieberman, P. M. Epigenetic Regulation of Kaposi's Sarcoma-Associated Herpesvirus Latency by Virus-Encoded MicroRNAs That Target Rta and the Cellular Rbl2-DNMT Pathway. *J Virol* **84**, 2697 (2010).

87. Qian, K. *et al.* Identification and Validation of Human Papillomavirus Encoded microRNAs. *PLoS One* **8**, e70202 (2013).

88. Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science* **31**, 8–22 (2022).

89. Sénécal, V., Barat, C. & Tremblay, M. J. The delicate balance between neurotoxicity and neuroprotection in the context of HIV-1 infection. *Glia* **69**, 255–280 (2021).

90. Kharel, P., Balaratnam, S., Beals, N. & Basu, S. The role of RNA G-quadruplexes in human diseases and therapeutic strategies. (2019) doi:10.1002/wrna.1568.

91. Kumar, S., Choudhary, D., Patra, A., Bhavesh, N. S. & Vivekanandan, P. Analysis of G-quadruplexes upstream of herpesvirus miRNAs: evidence of G-quadruplex mediated regulation of KSHV miR-K12-1-9,11 cluster and HCMV miR-US33. *BMC Mol Cell Biol* **21**, 67 (2020).

92. Wang, S., Li, H., Lian, Z. & Deng, S. The Role of RNA Modification in HIV-1 Infection. *Int J Mol Sci* **23**, (2022).

93. Kotaja, N. *et al.* The chromatoid body of male germ cells: Similarity with processing bodies and presence of Dicer and microRNA pathway components. *Proc Natl Acad Sci U S A* **103**, 2647–2652 (2006).

94. Zeng, Q., Ou, L., Wang, W. & Guo, D. Y. Gastrin, Cholecystokinin, Signaling, and Biological Activities in Cellular Processes. *Front Endocrinol (Lausanne)* **11**, 112 (2020).

95. Tripathi, S. *et al.* The gastrin and cholecystokinin receptors mediated signaling network: a scaffold for data analysis and new hypotheses on regulatory mechanisms. *BMC Syst Biol* **9**, (2015).

96. Józefiak, A., Larska, M., Pomorska-Mól, M. & Ruszkowski, J. J. The IGF-1 Signaling Pathway in Viral Infections. *Viruses* **13**, (2021).

97. Nonne, N., Ameyar-Zazoua, M., Souidi, M. & Harel-Bellan, A. Tandem affinity purification of miRNA target mRNAs (TAP-Tar). *Nucleic Acids Res* **38**, e20 (2010).

98. Desrochers, G. F. *et al.* Profiling Kinase Activity during Hepatitis C Virus Replication Using a Wortmannin Probe. *ACS Infect Dis* **1**, 443–452 (2016).

99. Liu, T., Zhang, L., Joo, D. & Sun, S. C. NF-κB signaling in inflammation. *Signal Transduction and Targeted Therapy 2017 2:1* **2**, 1–9 (2017).

100. Ménager, M. M. & Littman, D. R. Actin Dynamics Regulates Dendritic Cell-Mediated Transfer of HIV-1 to T Cells. *Cell* **164**, 695–709 (2016).

101. Kim, Y., Pierce, C. M. & Robinson, L. A. Impact of viral presence in tumor on gene expression in non-small cell lung cancer. *BMC Cancer* **18**, 1–13 (2018).

102. Chauhan, A. Unperturbed Posttranscriptional Regulatory Rev Protein Function and HIV-1 Replication in Astrocytes. *PLoS One* **9**, 106910 (2014).

103. Pongoski, J., Asai, K. & Cochrane, A. Positive and Negative Modulation of Human Immunodeficiency Virus Type 1 Rev Function by cis and trans Regulators of Viral RNA Splicing. *J Virol* **76**, 5108 (2002).

104. Paz, S., Krainer, A. R. & Caputi, M. HIV-1 transcription is regulated by splicing factor SRSF1. *Nucleic Acids Res* **42**, 13812 (2014).

105. Jabara, H. H. *et al.* A missense mutation in TFRC, encoding transferrin receptor 1, causes combined immunodeficiency. *Nature Genetics 2015 48:1* **48**, 74–78 (2015).

106. Krishna, B. A., Humby, M. S., Miller, W. E. & O'Connor, C. M. Human cytomegalovirus G protein-coupled receptor US28 promotes latency by attenuating c-fos. *Proc Natl Acad Sci U S A* **116**, 1755–1764 (2019).

107. Raber, J. *et al.* Central nervous system expression of HIV-1 Gp120 activates the hypothalamic-pituitary-adrenal axis: evidence for involvement of NMDA receptors and nitric oxide synthase. *Virology* **226**, 362–373 (1996).

6. SUPPLEMENTARY INFORMATION

Table 6.1. List of HIV-1 miRNA-like molecules predicted targets.

HIV-encoded miRNA-like molecule	Prediction Algorithm	Predicted target genes
sncHIV1_1	Target Rank	CNBP, DCP1A, CCDC88A, CCNJ, CLTC, DOCK7, DPYSL2, DSE, DNAJC6, CCN2, CBFB, CSRNP2, CAMTA1, CALML4, CALM1, CHIC1, CTNND1, CELF2, CHD7, CRYAB, CCND1, CAPZA2, ATP7A, ASCL1, ARHGAP24, BAG3, BCL2L11, C21orf91, C1QTNF, ATP13A3, BNC2, ATXN2, ATP10B, ATP2B1, BTBD7, BCL2, ARHGEF26, ARHGAP31, ACVR2B, ALDH1A3, ADD3, ANP32E. ANTXR1, ADSS2, AAK1, ACTR2, EIF1AX, ELAVL4, ENTPD1, EPB41L3, ERMN, ESCO2, ETAA1, ETNK1, ETV, FAM120A, FAM168B, FBXO11, FCHO2, FGF7, FNDC3B, FNIP1, FOXJ2, FOXP1, GABRA4, GCOM1, GMPPB, GMPR, GPBP1L1, GPM6A, GPR137B, GRIA3, HNRNPL, HNRNPU, ILF3, ITCH, ITPR2, KCMF1, KCTD12, KDELR2, KHDRBS1, KIAA1549, KLHL9, LCOR, LHFPL2, LRP6, LRRTM2, MAP1B, MAP6, MAPK1, MB21D2, MEX3C, MIB1, MIER1, MLLT6, MME, MORF4L1, MORF4L2, MTPN, MYT1L, NAA16, NAPEPLD, NBR1, NCOA3, NECAB1, NEDD4, NFKB1, NHP2, NLGN1, NOVA1, NPTX2, NR2C2, NRBF2, NUDT4, OSBPL5, PABPC1, PABPC3, PACS2, PAIP2B, PDE11A, PDE3B, PDGFRA, PDPK1, PLAG1, POLR2M, POU3F2, PPP1R9A, PPP2R3A, PPP2R5E, PPP4R4, PRDM16, PRKD3, PTEN, PTER, RAB22A, RAB7A, RAP1A, RBM24, RCOR1, RNF111, RNF125, RNF165, RNF38, RP2, RPA1, RPGRIP1L, RPL15, SCD, SCN1A, SECISBP2L, SLAIN1, SLC12A5, SLC26A2, SMARCAD1, SOCS5, SOX11, SSH1, ST13, STK4, SUZ12, SYN2, TCIM, TEAD1, TENT4B, TERB2, TET3, TIMM8A, TMX1, TP53INP2, TWSG1, UBE2G1, UBE3A, UBP1, UBR5, VPS4A, YTHDC1, YY1, ZBTB18, ZBTB21, ZBTB8A, ZCCHC2, ZDHHC18, ZFHX4, ZFYVE21, ZNF148, ZNF528, ZNF592
	Target Scan	DMXL1, DNAJB14, DCP1A, CNBP, CDH3, DNM3, CALD1, CFAP97, CHD6, DYRK2, CETN2, DCTN4, CCND1, CALML4, CTNND1, CRYAB, CTAGE6, CPSF2, CD34, CX3CR1, CCDC88A, CAB39, CCN2, CCNJ, CALM1, DYNLT3, DSE, CCDC47 , DCUN1D1, DCAF5, CHIC1, CAT, CAVIN2, CYBB, CBFB, CSRNP2, CAMTA1 CAPZA2, CNEP1R1, CHD7, CAP2, CYP51A1, CELF2, DNAJC6, CLTC, CANX, COL19A1, CPNE8, DYNC1LI2, CDCA7, DDX1, BCAT1, ARHGAP31, BNC2, BTBD1, ATP10B, ATP7A, ARGLU1, BCL10, ATP2B1, C1QTNF3, ASCC3, BTBD7, C21orf91, ATP13A3, C3AR1, BCL2, ARHGAP24, ARHGAP5, ASCL1, C19orf12, BTF3L4, BAG3. C12orf76, BCL2L11, BCL11A, AKTIP, ALDH1A3, ADD3, ACVR2B, ACTR2, ADH1B, ADAMTS18, ADK, ABCG2, ANOS1, ANP32E, ANTXR1, ADSS2, ABRA, ACSL1, ACTC1, APOLD1, AGL, ARF6, EBF2, EFCAB7, EIF1AX, ENTPD1, EPB41L3, EPHA3, EPHA7, ERAP1, ERC2, ERMN, ETAA1, ETNK1, ETV1, EXOC5, EXTL2, F5, FAM110B, FAM111A, FAM120A, FAM13B, FAM174A, FANCF, FBXL20, FBXL3, FBXO11, FCGR3A, FCHO2, FDX1, FGF2, FGF5, FGF7, FMO1, FNDC3B, FNIP1, FOXJ2, FOXO3, FOXP1, FUBP3, G2E3, GABRA4, GAS2, GLS, GMPR, GNG12, GOSR1, GPAM, GPBP1L1, GPM6A, GPM6B, GPR137B, GRAMD2B, GRIA2, HAUS6. HBS1L, HDGFL3, HEATR5B, HERC1, HFM1, HIF1A, HNRNPDL, HNRNPU, HSPA12A, IDS, IKZF1, IL16, IL1R1, ILF3, INSL5, INTS2, IPMK, IQGAP2, ITCH, ITPR2, KCNA1, CTD12, KIAA1210, KIAA1549L, KIF14, KLHL24, KLHL9, KRT23, LAMB1, LCA5, LCLAT1, LCOR, LHFPL2, LRP11, LRP6, LRRTM2, LRRTM4, LTBP2, LYPLA1, MAGI1, MAP1B, MAP3K2, MAP6, MAPK1, MARCHF6, MATN3, MDH1, MED30, MEX3C, MFSD14A, MIB1, MIER1, MLLT6, MME, MMP2, MORF4L1, MORF4L2, MSL2, MTDH, MTPN, MTURN, MXD1, MYT1L, NAA16, NBR1, NCAM1, NCOA3, NECAB1, NEDD4, NEFL, NEGR1, NEUROD4, NF1, NFAT5, NHP2, NIPAL2, NLGN1, NOVA1, NR2C2, NRBF2, NTRK2, NUB1, NUDT4, NWD1, ORC5, OSBPL3, OSBPL5, OSBPL8, OXR1, PABPC1, PABPC3, PAFAH1B1, PAIP2B, PCDH17, PCSK7, PDE11A, PDE5A, PDGFRA, PDGFRB, PDPK1, PDS5B, PELI1, PELI2, PERP, PKD2, PLAG1, PLCXD3, POLI, POLR1F, POU3F2, PPIL1, PPP1R9A, PPP2R5E, PPP3CA, PPP4R4, PRDM16, PRKCH, PRKD3, PRPF38B, PTEN, PTER, PTGER3, PTPN4, PTPRM, PYCR2, RAB22A, RAB23, RAB7A, RABL3, RAP1A, RASA1, RBL2, RBM24, RBM45, RCCD1, RCOR1, RFK, RGS2, RIMKLB, RNF111, RNF125, RNF13, RNF165, RNF38, RP2, RPGRIP1L, RPL15, RUFY2, RYR2, SCARB2, SCGB1D1, SCN1A, SCN3A, SCN8A, SECISBP2L, SERINC3, SERPINB7, SFRP4, SI, SIKE1, SLAIN1, SLC12A5, SLC26A2, SLC26A7, SLC35G1, SLC38A1, SLC5A8, SLC7A11, SLFN11, SLITRK6, SLK, SMARCAD1, SMC4, SMCHD1, SMIM15, SNTB1, SOCS5, SOS2, SOX11, SPAG9, SPIN3, SREK1, SRFBP1, SSH1, ST13, ST6GAL1, STX17, SUZ12, SYN2, TAF12, TAF4B, TASOR, TBC1D9, TCIM, TEAD1, TENT4B, TERB2, TIMM10B, TIMM8A, TMEM200A, TMEM267, TMEM63C, TMEM67, TMEM70, TMX1, TOM1L1, TOMM40L, TP53INP2, TPD52, TRA2B, TRIM2, TRIM23, TRIM33, TWSG1, TXLNB, UBE2G1, UBE3A, UBE4A, UBE4B, UBL3, UBP1, USP10, USP24, VPS13B,

		VPS26C, VPS4A, VTA1, WAPL, WASL, WNT16, WWOX YIPF5, YY1, ZBTB18, ZBTB21, ZBTB33, ZCCHC2, ZDHHC18, ZDHHC21, ZFHX4, ZFYVE21, ZMYND11, ZNF148, ZNF43, ZNF518B, ZNF528, ZNF592
	miRDB	DNM3, CNBP, CCNJ, DDX1, CAVIN2, DCUN1D1, CTNND1, DYNC1LI2, DMXL1, CBFB, CLTC, DCP1A, CCN2, CRYAB, DNAJB14, DCAF5, DCTN4, CYBB, CPSF2, EBF2, CANX, CCDC88A, CALML4, CFAP97, CAT, CAMTA1, CX3CR1, CYP51A1, CHIC1, CHD7, CAP2, CD34, CCDC47, CPNE8, CHD6, CALD1, CETN2, DOCK7, COL19A1, CDH3, CDCA7, CNEP1R1, DYNLT3, DYRK2, CTAGE6, CAB39, DSE, CAPZA2, DPYSL2, BTF3L4, ATP2B1, ASCL1, BTBD1, C12orf76, BNC2, BAG3, ATP10B, ARHGEF26, ATXN2, ARGLU1, ARHGAP5, BCAT1, C21orf91, ATP7A, ASCC3, BCL11A, BCL10, ARHGAP24, C3AR1, C19orf12, AAK1, AKTIP, ADAMTS18, ADH1B, ADK, ADSS2, ANOS1, ABCG2, ANTXR1, ABRA, ACSL1, ACTC1, ACTR2, APOLD1, AGL, ARF6, ACVR2B, EFCAB7, ELAVL4, EPB41L3, EPHA3, EPHA7, ERAP1, ERC2, ERMN, ESCO2, ETAA1, ETNK1, ETV1, EXOC5, EXTL2, F5, FAM110B, FAM111A, FAM120A, FAM13B, FAM168B, FAM174A, FANCF, FBXL20, FBXL3, FBXO11, FCGR3A, FCHO2, FDX1, FGF2, FGF5, FMO1, FNDC3B, FNIP1, FOXJ2, FOXO3, FUBP3, G2E3, GABRA4, GAS2, GCOM1, GLS, GMPPB, GMPR, GNG12, GOSR1, GPAM, GPBP1L1, GPM6A, GPM6B, GPR137B, GRAMD2B, GRIA2, GRIA3, HAUS6, HBS1L, HDGFL3, HEATR5B, HERC1, HFM1, HIF1A, HNRNPDL, HNRNPL, HSPA12A, IDS, IKZF1, IL16, IL1R1, ILF3, INSL5, INTS2, IPMK, IQGAP2, ITCH, KCMF1, KCNA1, KCTD12, KDELR2, KHDRBS1, KIAA1210, KIAA1549, KIAA1549L, KIF14, KLHL24, KLHL9, KRT23, LAMB1, LCA5, LCLAT1, LHFPL2, LRP11, LRP6, LRRTM4, LTBP2, LYPLA1, MAGI1, MAP1B, MAP3K2, MAP6, MARCHF6, MATN3, MB21D2, MDH1, MED30, MEX3C, MFSD14A, MIB1, MLLT6, MMP2, MORF4L1, MORF4L2, MSL2, MTDH, MTPN, MTURN, MXD1, MYT1L, NAA16, NAPEPLD, NBR1, NCAM1, NECAB1, NEFL, NEGR1, NEUROD4, NF1, NFAT5, NFKB1, NHP2, NIPAL2, NPTX2, NTRK2, NUB1, NUDT4, NWD1, ORC5, OSBPL3, SBPL8, OXR1, PABPC3, PACS2, PAFAH1B1, PCDH17, PCSK7, PDE3B, PDE5A, PDGFRB, PDPK1, PDS5B, PELI1, PELI2, PERP, PKD2, PLCXD3, POLI, POLR1F, POLR2M, POU3F2, PPIL1, PPP1R9A, PPP2R3A, PPP3CA, PPP4R4, PRKCH, PRKD3, PRPF38B, PTER, PTGER3, PTPN4, PTPRM, PYCR2, RAB23, RAB7A, RABL3, RASA1, RBL2, RBM24, RBM45, RCCD1, RFK, RGS2, RIMKLB, RNF111, RNF13, RNF38, RP2, RPA1, RPGRIP1L, RUFY2, RYR2, SCARB2, SCD, SCGB1D1, SCN1A, SCN3A, SCN8A, SECISBP2L, SERINC3, SERPINB7, SFRP4, SI, SIKE1, SLAIN1, SLC12A5, SLC26A2, SLC26A7, SLC35G1, SLC38A1, SLC5A8, SLC7A11, SLFN11, SLITRK6, SLK, SMC4, SMCHD1, SMIM15, SNTB1, SOCS5, SOS2, SPAG9, SPIN3, SREK1, SRFBP1, ST6GAL1, STK4, STX17, SUZ12, TAF12, TAF4B, TASOR, TBC1D9, TCIM, TEAD1, TENT4B, TET3, TIMM10B, TIMM8A, TMEM200A, TMEM267, TMEM63C, TMEM67, TMEM70, TOM1L1, TOMM40L, TP53INP2, TPD52, TRA2B, TRIM2, TRIM23, TRIM33, TXLNB, UBE2G1, UBE3A, UBE4A, UBE4B, UBL3, UBP1, UBR5, USP10, USP24, VPS13B, VPS26C, VPS4A, VTA1, WAPL, WASL, WNT16, WWOX, YIPF5, YTHDC1, YY1, ZBTB33, ZBTB8A, ZCCHC2, ZDHHC21, ZFHX4, ZFYVE21, ZMYND11, ZNF148, ZNF43, ZNF518B
sncHIV1_2	Target Scan	ANKRD28, ANKRD50, ATXN1, BCL2, BEND4, DDX3X, DUSP16, ELAVL2, EMC4, ENPEP, GOLPH3, KPNA4, NEGR1, PDE4D, PDYN, RNF11, RPS6KA3, SCD, SH2B1, SLC15A2, SMAD9, SNAP25, SNAP91, SORT1, SOX11, SP3, SRSF1, STAG2, STRBP, TEAD1, TFRC, TNPO1, TP53INP1, TRIB2, TTC7A, UBR3, ZFX, ZRANB2
	Target Rank	AMOTL1, ANKRD50, ARL13B, ARL13B, ATXN1, BAG5, BEND4, BTBD10, C15orf48, CAMTA1, CNOT6, CNTNAP2, DDX3X, ELAVL2, EMC4, FAM241A, FMR1, FZD5, GOPC, HERC4, KCTD20, KITLG, MAN2A1, MSX1, NEGR1, NUDT21, PAPPA, PELI1, PLCXD3, POLR2C, PTAR1, PTBP3, PTGER3, RAB14, RAD21, RCHY1, REPS2, RGS13, RNF11, RNF125, SORT1, SOX11, SP3, SPACA1, STAG2, TIGAR, TMEM245, TMEM30A, TNPO1, TP53INP1, TPP2, TTC7A, UBR3, ZRANB2, ZW10
	miRDB	AMOTL1, ANKRD28, ANKRD50, ARL13B, ATXN1, BAG5, BCL2, BTBD10, C15orf48, CAMTA1, CNOT6, CNTNAP2, DDX3X, DUSP16, ELAVL2, EMC4, ENPEP, FAM241A, FMR1, FZD5, GOLPH3, GOPC, HERC4, KCTD20, KITLG, KPNA4, MAN2A1, MSX1, NUDT21, PAPPA, PDE4D, PDYN, PELI1, PLCXD3, POLR2C, PTAR1, PTBP3, PTGER3, RAB14, RAD21, RCHY1, REPS2, RGS13, RNF11, RNF125, RPS6KA3, SCD, SH2B1, SLC15A2, SMAD9, SNAP25, SNAP91, SORT1, SP3, SPACA1, SRSF1, STRBP, TEAD1, TFRC, TIGAR, TMEM245, TMEM30A, TNPO1, TP53INP1, TPP2, TRIB2, TTC7A, UBR3, ZFX, ZRANB2, ZW10

Table 6.2. List of HIV-2 miRNA-like molecule predicted targets.

sncHIV2_1	Target Scan	MECP2, ANKRD6, ATP6V1A, BET1L, VCPKMT, SPATC1L, CREBRF, CNOT6L, CREM, FAM118A, GPNMB, PCLAF, MLX, OLFML2A, TENT4B, PCDHGA1, PCDHGA10, PCDHGA11, PCDHGA12, PCDHGA2, PCDHGA3, PCDHGA4, PCDHGA5, PCDHGA6, PCDHGA7, PCDHGA8, PCDHGA9, PEX19, PIGM, PRCD, RYBP, SACM1L, SH2D4B, SKP1, SLC16A2, SSR3, TAF9B, ARID1A, ARID1B, ARL5A, BAZ2A, TMEM229B, CAMK2D, CCDC88C, COIL, RTL8C, RTL8B, NTM, KIT, MAPK1, POLI, RAP2B, SCN8A, SORCS3, SOX11, SSH2, STXBP5L, UNC13C, WDR37, XIAP, GATAD2A, JARID1C, SNCA
	Target Rank	SLC16A2, CNOT6L, BAZ2A, GPM6B, MECP2, INO80D, PRPF4B, TENT4B, IGF1R, ACVR2B, PITPNA, WDR37, SSH2, ANKRD6, PEX19, MAPK1, ZC3H12C, CTDSPL, TBL1X, SEMA5A, YWHAZ, PIGM, TFEC, AMOT, GDA, SSR3, SKP1, SOX11, MECP3, ANKRD7, ATP6V1A, BET1L, VCPKMT, SPATC1L, CREBRF, CREM, FAM118A, GPNMB, PCLAF
	miRDB	MLX, OLFML2A, TENT4B, PCDHGA1, PCDHGA10, PCDHGA11, PCDHGA12, PCDHGA2, PCDHGA3, PCDHGA4, PCDHGA5, PCDHGA6, PCDHGA7, PCDHGA8, PCDHGA9, PEX20, PIGM, PRCD, RYBP, SACM1L, SH2D4B, SKP2, SLC16A3, SSR4, TAF9B, ARID1A, ARID1B, ARL5A, BAZ2A, TMEM229B, CAMK2D, CCDC88C, COIL, RTL8C, RTL8B, NTM, KIT, MAPK2, POLI, RAP2B, SCN8A, SOX12, SSH3, STXBP5L, UNC13C, WDR38, XIAP, GATAD2A, JARID1C, SNCA, SLC16A3, CNOT6L, BAZ2A, GPM6B, MECP3, INO80D, PRPF4B, TENT4B, IGF1R, ACVR2B, PITPNA, WDR38, SSH3, ANKRD7, PEX20, MAPK2, ZC3H12C, CTDSPL, TBL1X, SEMA5A, YWHAZ, PIGM, TFEC, AMOT, GDA, SSR4, SKP2, SOX12, MECP4, ANKRD8, ATP6V1A, BET1L, VCPKMT, SPATC1L, CREBRF, CNOT6L, CREM, FAM118A, GPNMB, PCLAF, MLX, OLFML2A, TENT4B, PCDHGA1, PCDHGA10, PCDHGA11, PCDHGA12, PCDHGA2, PCDHGA3, PCDHGA4, PCDHGA5, PCDHGA6, PCDHGA7, PCDHGA8, PCDHGA9, PEX21, PIGM, PRCD, RYBP, SACM1L, SH2D4B, SKP3, SLC16A4, SSR5, TAF9B, ARID1A, ARID1B, ARL5A, BAZ2A, TMEM229B, CAMK2D, CCDC88C, COIL, RTL8C, RTL8B, NTM, KIT, MAPK3, POLI, RAP2B, SCN8A, SOX13, SSH4, STXBP5L, UNC13C, WDR39, XIAP, GATAD2A, JARID1C, SNCA, SORCS3, SLC16A4, CNOT6L, BAZ2A, GPM6B, MECP4, INO80D, PRPF4B, TENT4B, IGF1R, ACVR2B, PITPNA, WDR39, SSH4, ANKRD8, PEX21, MAPK3, ZC3H12C, CTDSPL, TBL1X, SEMA5A, YWHAZ, PIGM, TFEC, AMOT, GDA, SSR5, SKP3, SOX13, ATP6V1A, BET1L, VCPKMT, SPATC1L, CREBRF, CNOT6L, CREM, FAM118A, GPNMB, PCLAF, MLX, OLFML2A, TENT4B, PCDHGA1, PCDHGA10, PCDHGA11, PCDHGA12, PCDHGA2, PCDHGA3, PCDHGA4, PCDHGA5

Table 6.3. Gene set enrichment analysis for genes present in HIV-1 interactome identified as potential targets of sncHIV1_1

GO term	Description	Category
GO:0051412	Response to corticosterone	Biological process
GO:0051385	Response to mineralocorticoid	Biological process
GO:1990089	Response to nerve growth factor	Biological process
GO:1990090	Cellular response to nerve growth factor stimulus	Biological process
GO:0035265	Organ growth	Biological process
GO:0001654	Eye development	Biological process
GO:0150063	Visual system development	Biological process
GO:0051051	Negative regulation of transport	Biological process
GO:0007423	Sensory organ development	Biological process
GO:0045216	Cell junction organization	Biological process
GO:0043197	Dendritic spine	Cellular component
GO:0044309	Neuron spine	Cellular component
GO:0097447	Dendritic tree	Cellular component
GO:0036477	Somatodendritic compartment	Cellular component
GO:0043005	Neuron projection	Cellular component
GO:0008143	Poly(A) binding	Molecular Function
GO:0070717	Poly-purine tract binding	Molecular Function
GO:0003727	Single-stranded RNA binding	Molecular Function

Table 6.4. Gene set enrichment analysis for genes present in HIV-1 interactome identified as potential targets of sncHIV1_2

GO term	Description	Category
GO:0033592	RNA strand annealing activity	Molecular Function
GO:0034046	Poly(G) binding	Molecular Function
GO:0140666	Annealing activity	Molecular Function
GO:0035613	RNA stem-loop binding	Molecular Function
GO:0070717	Poly-purine tract binding	Molecular Function

Table 6.5. Gene set enrichment analysis for genes present in HIV-1 interactome identified as potential targets of sncHIV2_1

GO term	Description	Category
GO:0006305	DNA alkylation	Biological process
GO:0006306	DNA methylation	Biological process
GO:0033391	Chromatoid body	Cellular component
GO:0016442	RISC complex	Cellular component
GO:0031332	RNAi effector complex	Cellular component
GO:0098794	Postsynapse	Cellular component
GO:0000785	Chromatin	Cellular component