**Introduction:**
HIV/AIDS is the most widely spreading disease of the 21\textsuperscript{st} century. Its virulent and elusive nature comes from the targeted action of the virus against the immune system and its highly mutable genome. Current anti-retroviral therapies continually fail to completely eradicate the virus due to rapid viral mutations and a latent viral population integrated into the genome of host cells. Activation from latency is a poorly understood process known to require the action of several host proteins. Recent technological advancements have independently provided a means to (1) identify sites of HIV integration on a large scale using massively parallel pyro-sequencing\textsuperscript{3} and (2) identify host proteins or important RNAs which contribute to HIV replication using large-scale siRNA screens\textsuperscript{8}. I have proposed a mechanism combining these technologies to identify RNAs or RNA products involved in activation of latently integrated HIV.

**Background:**
The startling incidence of HIV/AIDS has raised its status to a modern epidemic: in 2006, an estimated 39.5 million people worldwide were affected with HIV/AIDS, including 4.3 million new cases and 2.9 million AIDS-related deaths\textsuperscript{1}. HIV-1, the virus that causes AIDS, is a highly elusive retrovirus that assaults the immune system by targeted (host-receptor mediated) attack of \textit{CD4+} cells. As the virus encodes only 15 genes, it must harness the host machinery to propagate. HIV infected cells evade detection by the body by forming an envelope of the cells' own membrane and through protective glycosylation. Additionally, because HIV is a retrovirus, it has a high rate of mutation and replication, and thus adaptation. Upon infection, a portion of the HIV population actively replicates in the body. However, the virus also establishes an inactive “reservoir” population by stably and dormantly integrating its own genome into the genome of otherwise healthy established cell populations, commonly in lymph nodes, bone marrow and other macrophages (HIV mechanism reviewed in \textsuperscript{2}). This latent population is often a result of unsuccessful host-genome integration, where the virus has inserted itself into a heterochromatic portion of the DNA\textsuperscript{3}. Although developments in
**highly active anti-retroviral therapy** (HAART) are able to minimize the levels of actively replicating virus in the body (viral load) below detection levels, spontaneously activated latent reservoirs of HIV make the virus impossible to destroy.

Treatment with the non-tumorigenic, phorbol-ester *Prostratin* and the HDAC1 inhibitor *viproic acid* can generally induce nearly 80% of the latent HIV population to activate *in vivo*, rendering the newly activated viral population susceptible to HAART\(^4\). However, in spite of this treatment, the residual latent and low-level undetectable viral populations continue to replicate\(^5\). With each replication cycle the virus undergoes increasing opportunity for mutation and inevitably becomes resistant to all known HAART, chronically and progressively leading to symptomatic HIV, AIDS and AIDS-related fatality. Indeed, a means of elimination of the latent population could potentially extend patient lifespan and is necessary (though not likely sufficient) for complete viral eradication from the body.

**Preliminary Data of Supporting Techniques**

Two independent studies have recently shaped the HIV field. Most notably, the use of large-scale siRNA screens has been used to identify nearly 270 host proteins whose knockdown positively affects the fitness of HIV-infected cells.\(^6\) As in Brass *et al.* (2008) cell populations were infected with HIV and individually treated with an siRNA from a human library of approximately 40,000 siRNAs, generating a knockdown of a single host protein in each population. HIV replication potential was then assayed by relative host cell survival and fitness. Namely, if the knockdown of a host protein yielded increased host cell survival and proliferation, that protein was implicated in furthering HIV propagation. A shortcoming of this assay is that the specific role of each identified protein and its relationship to HIV propagation are left to be elucidated. Additionally, the study did not select for latent infections, so the experiment reflected the dynamics of an actively replicating HIV population much more than a latently infected population. Whether or not latent and active infections require the same proteins to replicate remains largely unknown.
A second advancement utilizes massively parallel pyrosequencing (MPSS) to locate HIV integration sites on a high-throughput, genome-wide scale. Briefly, in Wang et al. (2007) Jurkat cells were infected with a pool of an HIV-1 strain carrying GFP (and the full HIV-1 genome). Fifty independent such infections took place. Cells were harvested 72 hours post infection and an 80% infected population was confirmed by FACS profiling. The genomic DNA was then extracted from these cells and cut with restriction enzymes, ligated to linkers, amplified by nested PCR and sequenced by MPSS. Fragments which contained 3 bp of the long terminal repeat regions (endpoints) of the virus and also had 98% sequence homology were positively identified as a site of integration. Once again, this technique can only significantly describe sites of active integration.

Research Design and Methods:
The specific aims of this study are outlined and then explained below, as well as in Figures 1 and 2:

**Aim 1:** *Generation of a polyclonal latently HIV-1 infected cell population*

**Aim 2:** *Identification of integration sites across the polyclonal cellular populations of Aim 1*

**Aim 3:** *RNAi screen to identify RNAs important for latent HIV activation*

**Aim 4:** *Low throughput verification of targets*

**Aim 1: Generation of latently infected cells (Figure 1)**
To generate a polyclonal population of latently infected cells, Jurkat cells will be infected *en masse* (high MOI) with an HIV-1 population carrying GFP as in Wang et al. (2007). The cells will then undergo a 72 hour incubation, allowing the virus ample time to integrate into the host genome. A 90 percent [active] infection rate will be verified by FACS profiling using the GFP tag. Susceptibility of the virus to a variety of cocktails will be tested on a small population of GFP-verified, actively infected Jurkat cells to determine an effective form of HAART. The large pool of infected cells will then be treated with this HAART, inhibiting HIV-1 replication and causing actively infected cells to die. The incubation in anti-HIV cocktails will thus select for latently infected and uninfected cells. An approximate ratio of uninfected to infected cells can come from qPCR against the viral genome compared to qPCR of another gene, such as actin. Primer
efficiency remains questionable but can be loosely gauged by comparison to the FACS-verified actively infected population where viral load can also be estimated. Since latent infection is much rarer than active infection, the entire process (from infection onwards) may be repeated on a population until qPCR reveals that 80 percent of cells are latently infected. This will serve as a polyclonal latently infected population. While this process may not be extremely efficient, it is important to note, firstly, that HIV will infect uninfected cells rather than latently infected cells\(^2,5\), improving efficiency of re-infection and preventing duplicate infections, and that only small pool volumes are required for future experiments.

**Aim 2: MPSS to determine integration sites**
The above polyclonal, latently infected population will undergo MPSS as performed in Wang *et al.* (2007) to determine the distribution of HIV integration in *latently* infected cells. This is an important step because one must be able to rule out that any findings reported on polyclonal populations are artifacts of integration site bias.

**Aim 3: RNAi screen to ID proteins necessary for latency activation (Table 1)**
Applying the technique of Brass *et al.* (2008) I will perform independent genome-wide RNAi screens on the monoclonal populations to identify RNAs necessary for HIV propagation. Because the cells will be nearly all latently infected, *increased* cell death as well as expression of GFP will be used to identify active HIV replication. The knockdown of an RNA that induces HIV replication (as compared to uninfected cells given the same treatment) is a result of activation of the latently inserted viral genome. These RNAs will be identified as important for suppressing latent HIV activation.

Next I will perform an siRNA screen simultaneously treating the cells with *Prostratin* and viproic acid. *Prostratin* and treatment with viproic acid have been shown to cause widespread HIV activation from latently infected populations. This time, the opposite phenotype: *decreased* cell death relative to the former screen or to uninfected cells given the same treatment will be assessed. In essence, the knockdowns in which viral replication (ie cell death and GFP fluorescence) are *inhibited or reduced* will identify RNAs important for activation of latent HIV populations.
The RNAi screens require sufficient controls. Firstly, one must perform the screen on uninfected Jurkat cells with and without drug to gauge a relative amount of cell death or survival from the knockdown and/or drug alone. Additionally, there is a possibility that an actively integrated virus has mutated to become resistant to or has evaded elimination by the given HAART. For this reason, one should assay GFP fluorescence of each well before and after knockdowns to ensure the virus is not actively replicating before the knockdown. Lastly, because the cells are polyclonal, it is necessary to perform several biological replicates. Further verification of targets will be explored in Aim 4.

**Aim 4: Verification**

After identifying the siRNAs important for viral activation, one should repeat the screen on larger pools to confirm that the RNAs are indeed important for viral activation. For positive candidates, MPSS (as in Aim 2) of the resulting cell populations should be performed to identify if the actions of the siRNA affects (on a broad scale) certain integration sites. The screen in Aim 2 will eliminate biases due to an initial distribution. Specifically one could combine this data with prior data in the literature to further characterize the epigenetic landscape or nucleosome positioning of latent integration sites. Using this information one could hypothesize whether the activations due to the knockdown were the result of general processes, such as increased accessibility due to widespread heterochromatic activation, or if the RNA or RNA product was part of a mechanism specific to HIV.

**Summary**

I have proposed a mechanism to identify host proteins implicated in latent HIV activation using massively parallel pyrosequencing and high throughput siRNA library screening. The results of this study may identify drug targets to eliminate latent viral populations and are a step in the direction towards viral eradication in the body.
References

1) World Health Organization
3) Wang et al., Gen Res 17, 1186 (2007)
6) Brass et al., Science 319, 921 (2008)