Control of HIV-1C Replication Both in Vivo and in Vitro

Control of HIV-1C Replication Both In Vivo and In Vitro

A dissertation presented

by

Wen Xie

to

The Committee on Higher Degrees in Biological Sciences in Public Health

in Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological Sciences in Public Health

Harvard University

Cambridge, Massachusetts

Aug, 2014

Dissertation Advisor: Professor Max Essex Wen Xie

Control of HIV-1C replication both in vivo and in vitro

Abstract

In this study, we established an HIV latency model using pluripotent stem cells *in vitro*. We found that a HIV accessory protein *nef* is responsible for full-length HIV silencing in human pluripotent stem cells. This effect is mediated by host factors NMT1, NMT2, Hck, PAK2, ACOT8. We further suggested that deprivation of host factors can only partially reactivate the virus.

We performed a genome-wide association study (GWAS) on Botswana treatment-naïve AIDS patients infected by HIV-1C. By applying a statistical method using functional principal component analysis to approximate the underlying trajectories of longitudinal CD4 and VL, we showed three SNPs mapped to genes of HCG22, ZBTB7C and CCNG1 are significantly associated with AIDS disease progression.

Our study provided insight in HIV latency mechanisms *in vitro* and also suggested new mechanisms of host controlling viral replication in African genetic backgrounds. Our result is important for both understanding the pathogenesis of HIV as well as treatment prioritization for AIDS intervention.

Preface

While this thesis will present the results of my academic work in graduate school, I first want to step back and reflect on those who have given me the most crucial support as I pursued my dream. I would like to thank Dr. Max Essex, my thesis advisor, for his support of all aspects of my PhD. His accomplishments have given me the strongest encouragement every day to become a scientist who, like him, significantly impacts both the basic science of HIV and global AIDS public health interventions. I also want to thank Dr. Tun-Hou Lee, my PhD supervisor. Since my first day at Harvard, he has mentored me in every aspect of my study. I want to acknowledge Dr. George Daley, my thesis co-supervisor. I will never regret joining his lab, as it is the place which truly opened my mind in science and showed me the joy of being a scientist. I want to thank Dr. Tianxi Cai, with whom I closely collaborated on my GWAS project. I feel privileged to have been their student!

I also want to thank many people who provided me advice or collaborated with me during my PhD program: Drs. Richard Marlink, Marianne Wessling-Resnick, Marc Lipsitch, Laurie Glimcher, Barry Bloom, Oliver Hoffmann and Rory Kirchner.

I also gratefully acknowledge my Dissertation Defense Committee, Drs. George Seage, Stephen J. O'Brien and Monty Montano, for offering me their expertise and guidance. And finally, I thank my family members and friends for their great support for me during my PhD program—especially my mother, Lijian Wang, my father, Jingyan Xie, and my fiancé, Yusheng Qi.

TABLE OF CONTENTS

LIST OF FIGURES WITH CAPTIONS

LIST OF TABLES WITH CAPTIONS

Summary

After 33 years since the recognition of acquired immunodeficiency syndrome (AIDS) in 1981, there are still no effective therapies to cure AIDS. HAART (highly active antiretroviral therapy) interruption results in a recurrence of viremia requiring life-long adherence to this therapy. Interestingly, without antiviral drug treatment, AIDS patients also show variation in the disease progression to AIDS and clinical outcome after infection with Human immunodeficiency virus (HIV) (Fellay J, 2007) and a portion of patients are able to maintain low viremia for a long time without treatment. These phenomena suggest that both host factors and viral factors play important roles in controlling the viremia which leads to long-term HIV latency. Understanding how HIV establishes its latency and why people display different levels of HIV viremia control is important for both developing an effective treatment to eradicate the HIV long-term reservoir, as well as prioritizing patients predisposed to early development of AIDS for initiation of treatment.

Due to the lack of an efficient HIV latency cell model *in vitro*, we established an HIV latency model using pluripotent stem cells. We found that full-length HIV defective of envelope gene (HIV-Δ*env*) is silenced in human pluripotent stem cells but not other differentiated lineages (fibroblast, blood). By knocking out a HIV accessory protein Nef or supplementing Nef into Nef defective HIV infection (HIV-Δ*env*/*nef*), we demonstrated that Nef is responsible for silencing the virus at an epigenetically transcriptional level by recruiting histone modification players to perform methylation at H3K9 and

demethylation at H3K4 on the HIV genome. This effect is independent of viral integration site selection. Host factors (NMT1, NMT2, Hck, PAK2, ACOT8) that have been reported to interact with Nef are also highly involved in establishing the latency. We further suggested that Nef leads to long-term HIV silencing, and deprivation of host factors can only partially reactivate the virus. To further test the hypothesis that Nef and its associated host factors are responsible for HIV latency *in vivo*, we need to replicate the experiment in natural HIV reservoirs such as hematopoietic stem cells (HSCs).

To study HIV latency and host genetic restriction to viral replication *in vivo*, we performed a genome-wide association study (GWAS) on Botswana treatment-naïve AIDS patients. We applied a statistical method using functional principal component analysis to approximate the underlying trajectories of longitudinal CD4 and VL. By conducting an initial GWAS analysis on a Tswana HIV-1-Subtype-C cohort sized 317 and replicating it in an independent Tswana HIV-1C cohort sized 239 , we showed three SNPs mapped to genes of HCG22, ZBTB7C and CCNG1 are significantly associated with AIDS disease progression. Adjusting for age and sex does not change our result. Our result suggested that HLA mediated HIV restriction plays an important role in viral control in African genetic backgrounds, and we revealed that the p53/p21 pathway may be an important element in viral replication restriction *in vivo*. Exon sequencing and *in vitro* knockdown assays may be needed to further analyze the causal relationship.

In summary, by establishing a new *in vitro* model of HIV latency using human pluripotent stem cells, we demonstrated that HIV-*nef*, together with both pluripotency

associated cellular factors and factors that interact with Nef, may contribute to establishing the HIV epigenetic silencing *in vitro*. By performing a GWAS on an HIV-1C African population, we further identified three host genes that may contribute to establishing long-term HIV host control in treatment naïve patients. Our *in vitro* study raised important host and viral targets that can be drug targets to reactivate the HIV latent reservoir for patients on HAART, while our GWAS study provided insight in mechanisms of HIV host restriction *in vivo*, which is important for both HIV vaccine design and treatment prioritization.

Chapter I Introduction

I.1 Overview of AIDS pandemic

a. Discovery of HIV and early studies on host-viral interaction

The year of 2011 marked the thirtieth year of the AIDS pandemic. In 1981, the world first recognized the existence of AIDS when most patients with AIDS died within 2 years before the discovery that AIDS was caused by a new immunodeficiency virus know as HIV-1(Rothenberg et al., 1987; Pomerantz and Horn, 2003). The first big breakthrough in AIDS happened in 1983 when scientists first identified viruses associated with AIDS(Barre-Sinoussi et al., 1983)and subsequently discovered AIDS was caused by a retrovirus HIV-1(Gallo et al., 1984; Popovic et al., 1984). The methodology used to link animal and human retroviruses to immunosuppressive disease led to the development and implementation of a blood-screening antibody test indicating the presence or absence of HIV-1 (Kitchen et al., 1984; Lee et al., 1984a; Sarngadharan et al., 1984; Allan et al., 1985b; Silberner, 1985; Archibald et al., 1986). The understanding of the virus rapidly progressed as scientists decoded the functional units of the HIV genome(Lee et al., 1984b; Wain-Hobson et al., 1985; Wong-Staal et al., 1985; Haseltine and Patarca, 1986; Lee et al., 1986). The discovery of specific host immune response to a particular region on viral genome (gp120/160) was another big step towards understanding of host-viral interaction that later on greatly facilitated the development of anti-viral therapy and vaccines(Allan et al., 1985a; Barin et al., 1985).

b. Current AIDS pandemic

In 2012, UNAIDS estimated that 35.3 (32.2–38.8) million people were living with HIV. There were 2.3 (1.9–2.7) million new HIV infections globally, showing a 33% decline in the number of new infections from 3.4 (3.1–3.7) million in 2001(UNAIDS, 2013b). At the same time the number of AIDS deaths is also declining in that there were 1.6 (1.4–1.9) million AIDS deaths in 2012 while there were 2.3 (2.1–2.6) million in 2005(UNAIDS, 2013b).

However, diversity of the AIDS pandemic among countries continue to persist. Sub-Saharan African counties are still the most affected, followed by Eastern Europe and the Caribbean (Figure 1).The AIDS pandemic is especially high in southern Africa with an overall prevalence among adults of up to 31% in Swaziland, 25% in Botswana, and 17% in South Africa(UNAIDS, 2013b). Despite the fact that global investment in the AIDS response jumped from US$3.8 billion in 2002 to US$18.9billion in 2012(Zablotska, 2013), countries with less sustained investments and less effective systems to deliver antiretrovirals to women and infant feeding-based prevention programmes still experience an increase in AIDS infection(Kuehn, 2006), with 50% of eligible patients not receiving HAART(UNAIDS, 2012). It is worth noticing that there are two common features of those countries: first, the HIV-1 subtype C infection has become hyperendemic in those countries rather than subtype B; second, the average income of those countries are low. These two facts pointed to the necessity to prioritize treatment resources to patients more susceptible to AIDS development and transmission based on the knowledge of host genetic restriction to HIV, especially the African population infected by HIV subtype C.

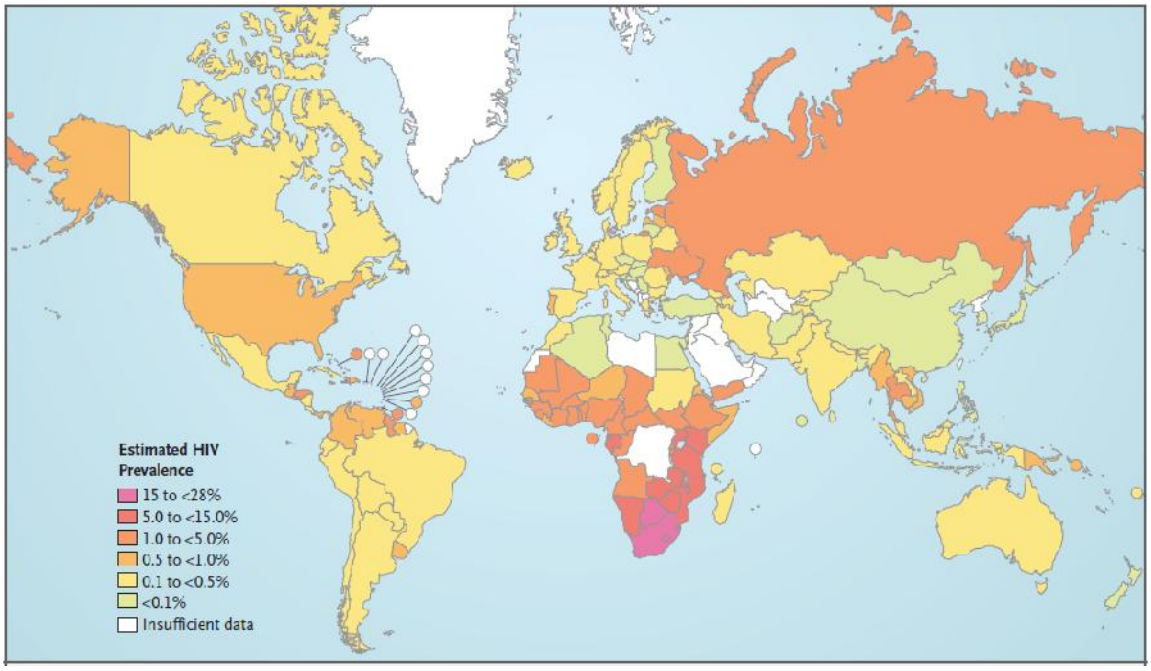

Figure 1. World Map of Prevalence of HIV Infection 2011.
Data are from UNAIDS(2012),UNICEF (www.unicef.org), and the World Bank (www.worldbank.org). The figure is adapted from reference(Piot and Quinn, 2013).

I.2 Current research on HIV/AIDS intervention

Research of HIV/AIDS intervention has focused on four main steps: prevention of transmission and diagnostics, disease progression monitoring before HAART, effectiveness of HAART treatment, and opportunistic outcomes (Figure 2).

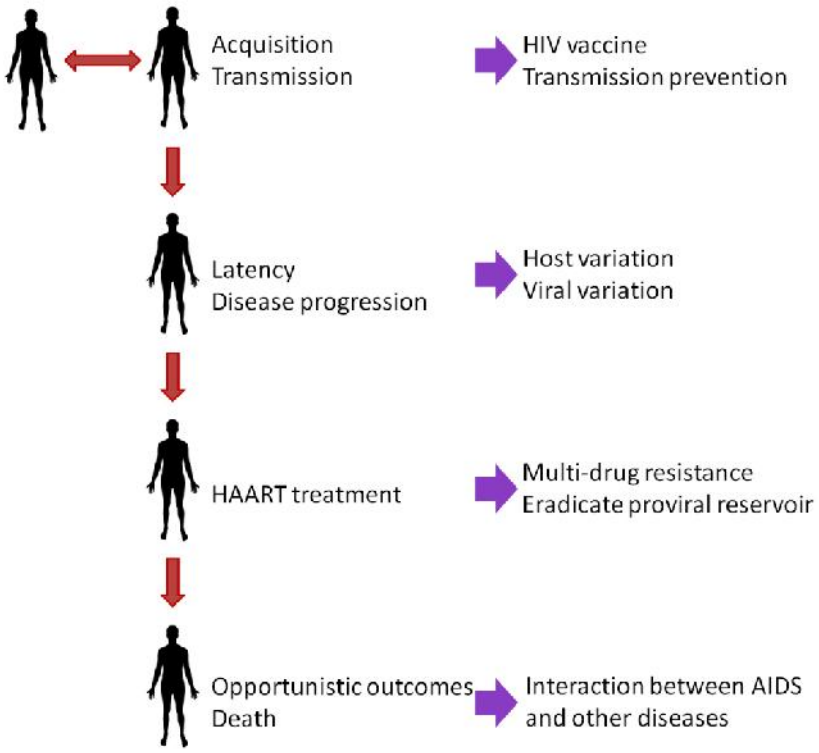

Figure 2 Four major areas of HIV/AIDS research

a. HIV transmission prevention

In the area of acquisition/transmission prevention, the research topics include developing an effective HIV vaccine and effective methods to prevent viral transmission including mother-to-child transmission and transmission through sexual activities. Recent studies showed that "treatment as prevention" as a new and promising direction (Cohen et al., 2011a; Padian et al., 2012; Shapiro et al., 2013).

Another area in HIV transmission prevention research is development of an HIV vaccine. Despite 30 years of effort, there is still no effective HIV-1 vaccine. The HIV-1 infection life cycle begins with host cell recognition followed by entry mediated by CD4 and CCR5 or CXCR4 on the host cell surface with the HIV-1 envelope on HIV virion(Figure

3). Studies suggested that during acute infection with HIV-1, serum viral loads peak approximately 3 weeks after transmission and then fall to a set point determined by the host immune system (Daar et al., 1991; Cohen et al., 2011b). Importantly, people have different viral load set point levels suggesting host genes may play a role in viral load control. Previous GWAS studies as well as other genetic research suggested that in rare individuals who carry specific HLA alleles such as HLA-B*5701, T cell responses can even reduce viremia to undetectable levels for prolonged periods of time(Migueles et al., 2000). This result raised the importance of T cell responses for partial control of viral replication in the early phases of HIV-1 infection. However, HLA-B*5701 protective allele can only explain less than 10% of the variation(Fellay et al., 2007), indicating other mechanisms involved in natural host control of HIV. Previously it has also been shown that patients with a homozygous defective CCR5 gene (CCRΔ32) are resistant to HIV infection(Huang et al., 1996; O'Brien et al., 1997; Quillent et al., 1998). Understanding the host genetic restriction to viral replication will be essential for developing a successful HIV vaccine.

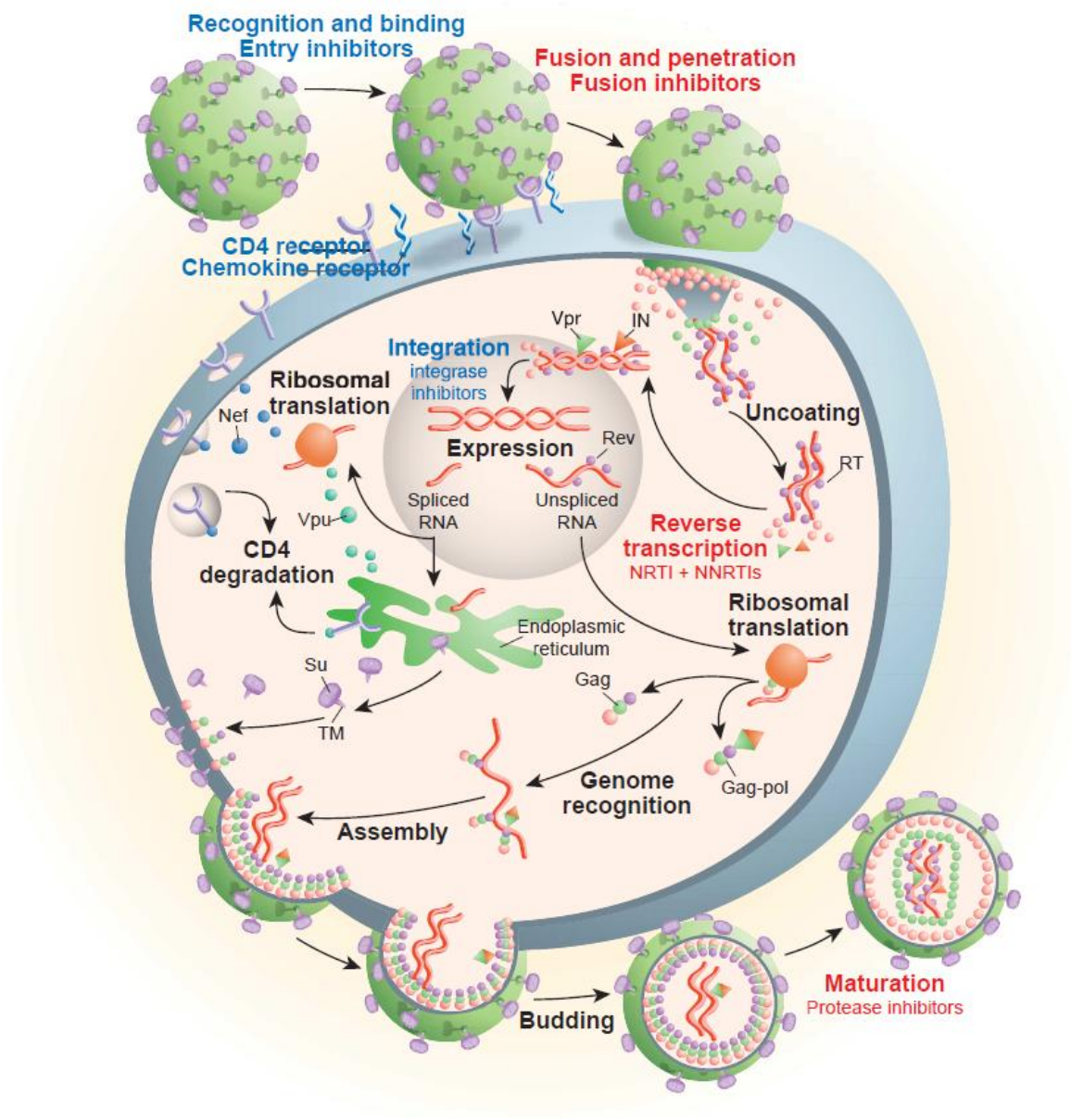

Figure 3HIV-1 life cycle.
The virus is recognized by host CD4 and CCR5/CXCR4 on the surface of cells and after binding and fusing to the host cell, the virion core enters the cell. Reverse transcriptase performs reverse transcription that synthesize the proviral double-stranded DNA, which will be transported to the nucleus and integrated into the host genome. The viral genome is then transcribed and the virus proteins are translated using host ribosomes. New virions are created by assembly and budding through the infected cell membrane. The figure is adapted from reference (Pomerantz and Horn, 2003)

b. Before HAART

The most heterogeneity in disease progression was displayed before patients started HAART. The disease course of AIDS is shown in Figure 4. However, there is considerable heterogeneity in virus control and progression rates that cannot be fully explained by environmental or viral factors(An and Winkler, 2010). These facts lead to the hypothesis that the heterogeneity of host genetics are responsible for explaining some variation in AIDS disease progression. The extreme phenotypes that deviate from the expected response to HIV exposure can be categorized into the following three kinds(An and Winkler, 2010):

(i) exposed uninfected (EU) individuals who are resistant to HIV acquisition even after multiple, high risk exposures (O'Brien and Nelson, 2004; Shacklett, 2006)

(ii) long-term non-progressors(LTNP) who maintain stable CD4 levels and low virus load for ten or more years(Kumar, 2013)or elite controllers (EC) (1% of HIV-infected patients), who are infected with replication-competent virus but control HIV replication to <50 copies/ml(Walker, 2007; Miura et al., 2008).

(iii) fast progressors who cannot control viremia and develop AIDS within three years of infection (O'Brien and Nelson, 2004)

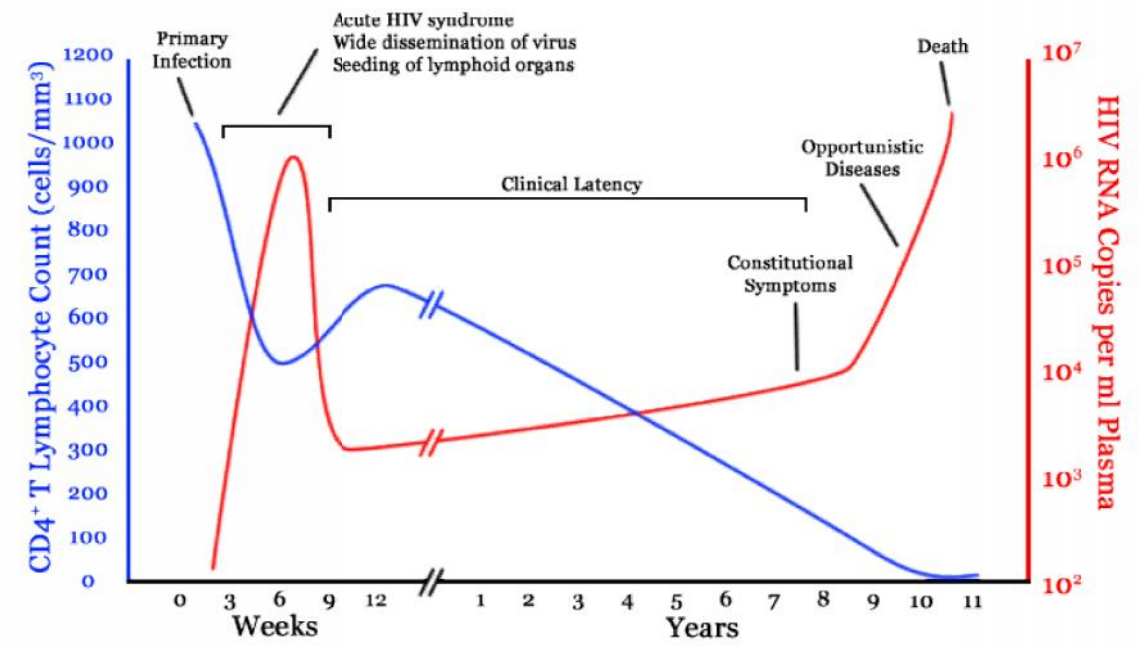

Figure 4. A typical natural course of HIV infection.
Acute phase lasts for 6–12 weeks with flu-like symptoms, peak virus load and drop in CD4+ T cells. Chronic asymptomatic phase lasting on average 7–10 years; following the acute phase virus replication reaches a steady level known as the 'set point'. This is followed by AIDS onset associated with increasing virus replication and declining $CD4_+$ cell counts to $<200/mm_3$; time to AIDS onset varies between individuals from as little as 2 years to more than 15 years after seroconversion. Figure modified from references (Fauci et al., 1996; An and Winkler, 2010)

c. Highly Active Antiretroviral therapy (HAART)

After the patient CD4 drops to a certain threshold (typically 200-350/μl), they are eligible for HAART treatment(Dybul et al., 2002). Despite the development of multidrug resistance (Bangsberg et al., 2000; Hirsch et al., 2000), patients cannot be cured by HAART due to the viral latency. "Retroviral latency" is defined as integrated replication competent provirus with no active transcription(Pomerantz, 2002a). Multiple host genetic factors are involved in silencing the proviral genome reviewed in I.4b. However, upon suspension of HAART, latent proviral genes get reactivated and the mechanism of viral reactivation and the variation of the timing of viral reactivation is largely unknown. Studies of the mechanism of host genetic control of the HIV proviral genome using a

good *in vitro* cell model is needed to understand the HIV reactivation process and to understand possible mechanisms for the eradication of the proviral reservoir.

I.3 Challenge of curing AIDS

a. Proviral reservoir

As discussed above, HAART can substantially reduce the viral load within weeks of initiation. However, upon stopping HAART, the viral remission can last for only a short period and virus will eventually reactivate through unknown mechanisms(Saez-Cirion et al., 2013). Even after early initiation of HAART in adults, only 5-15% could potentially lead to prolonged remission while the rest experience rapid viral load rebound after stopping HAART(Saez-Cirion et al., 2013). In the case of the famous "Mississippi baby", an HIV-infected baby who received HAART within 30 h of delivery, stayed on HAART for 18 months and seemed to be cured(Persaud et al., 2013; Shiau and Kuhn, 2014). However after 27 months off HAART, the viral load rebounded(NIAID, 2014). This fact that HIV can remain hidden for years and eventually rebound implicated that host control of HIV provirus is limited even at very early initiation of HAART and remission of viral replication.

Latent replication-competent HIV-1 provirus exists mainly in resting memory (CD45RO) CD4+ T lymphocytes(Pomerantz, 2002b). Yet it has been suggested that other cell types such as monocytes and resting naive (CD45RA) CD4+ T lymphocytes can also serve as proviral reservoirs in patients receiving virally suppressive HAART(further reviewed in II.2)(Sonza et al., 2001). Naive CD4+ T lymphocytes are rarely directly infected,

therefore the naive CD4+T lymphocytes harboring HIV proviral genes were reported to be generated by means of reversion from memory CD4+ T cells (Pierson et al., 2000).

It has been suggested that the proviral reservoir can also serving as a source for documentation of viral mutations. Viruses isolated from resting CD4+ T lymphocytes from the peripheral blood of patients with undetectable levels of viral RNA in plasma may harbor resistance mutations in the reverse-transcriptase(RT) and protease (PR) genes of HIV-1(Wong et al., 1997).

Outside the peripheral blood, lymphoid tissues can also bear latent HIV-1 proviral genes. It has been suggested that gut-associated lymphoid tissue also appears to be an important site for early HIV-1 replication in these mucosal lymphoid tissues, thus serving as an important reservoir of HIV-1 outside of the peripheral blood for HIV-1(Veazey et al., 1998). In addition to lymphoid tissues, tubular epithelial
cells of the kidney has also been reported as another potential reservoir for HIV-1(Winston et al., 2001).

b. Lack of effective biomarker for host-viral interaction

Another challenge in curing HIV is the lack of a reliable biomarker for the total body burden of replication-competent HIV that persists in people on HAART. As virus may develop antidrug mutations, without an effective biomarker, it is hard to monitor the burden and reactivation of provirus from reservoirs in the patients after long term

antiretroviral treatment. The "Mississippi baby"(Persaud D, 2000; Persaud et al., 2013) had a viral rebound after 27 months of HAART suspension, and two patients infected with HIV in Boston, MA, USA who received a stem-cell transplant for lymphoma(Henrich et al., 2013) and had virus rebound 12 and 32 weeks after stopping HAART. In all these three cases, the currently available assays failed to detect evidence of HIV return. These tragedies indicated the importance of identifying a robust biomarker that will predict time to viral rebound once HAART is stopped.

I.4 Host restriction to HIV in vitro

a. Host restriction at pre-integration level

Many detailed molecular mechanisms of retrovirus or lentivirus latency have reported that both the host factors and viral factors are highly involved. Reported host pre-integration restriction factors include APOBEC3G that inhibits HIV-Δvif by digesting the reverse transcriptional machinery(Huthoff H, 2008), Trim5α that interacts with capsids to prohibit early entry viral-life cycle (Stremlau M, 2004; Huthoff H, 2008) and p21which interacts with integrase(Zhang J, 2007). Mutations in the cellular proteins that facilitate viral entry such as CCR5-Δ32(Loetscher P, 2000) and integrations such as HMG1(Farnet CM, 1997), LEDGF(Llano M, 2006; Shun MC, 2007b), emerin(Jacque JM, 2006; Shun MC, 2007a), microtubules(McDonald D, 2002) and PIC can also lead to HIV latency in both field research and cell models.

b. Host restriction at post-integration level

For post-integration mechanisms, host microRNAs (miR-28, miR-125b, miR-150, miR-223 and miR-382) (Omoto S, 2004; Han Y, 2007a; Huang J, 2007) were shown to affect HIV mRNA stabilities. However, most mechanisms of post-integration latency operate at the transcriptional level(Mayte Coiras, 2009). Proviral integration site and orientation could both affect the expression of proviral genes(Mayte Coiras, 2009; Shan L, 2011), though HIV-1 tends to insert in host transcriptionally active sites (Schröder AR, 2002) and most inducible HIV-1 proviruses can be found in intronic regions of highly expressed host genes(Han Y, 2004; Lewinski MK, 2009).The orientation and position that the HIV viruses integrate into the host genome affects transcription. Insertion sites that are divergent, convergent or parallel relative to a proximate host promoter could result in transcriptional interference(Callen BP, 2004; Crampton N, 2006; Han Y, 2008; Lenasi T, 2008; Lewinski MK, 2009). These findings indicate that chromosomal environment at the site of integration and the accessibility of host and viral transcriptional factors have substantial influence on viral latency.

Epigenetic silencing is a major mechanism of retroviral transcriptional silencing and histone deaceylation and histone methylation are both highly involved in the HIV model of control in Hela cells (du Chéné I, 2007; Tyagi M, 2007), resting T cells(Archin NM, 2009), Jurkat cells(Tyagi M, 2007; Pearson R, 2008; Friedman J, 2011), the murine leukemia virus (MLV) model of mouse embryonic stem cells (mES) (Wolf D, 2009) and the L1 retrotransposon in human embryonic carcinoma cells(Garcia-Perez JL, 2010). Late SV40 factor(Ylisastigui L, 2004), Yin-yang1(YY1)(Gordon S, 2006), C-promoter binding factor 1(CBF-1)(Tyagi M, 2007) and chicken ovalbumin upstream promoter-

transcription factor-interacting protein 2 (BCL-11B) (Marban C, 2007) can bind to HIV LTR and recruit HDACs which will suppress transcription, while *Tat* can recruit HAT proteins such as CREB binding protein (CBP), CBP-associated factor (PCAF) and human general control of amino acid synthesis protein 5(hGCN5) to hyper-acytelate 5' LTR and lead to viral expression(du Chéné I, 2007). Histone-lysine N-methyltransferase suppression of variegation 3-9 homologue 1 (SUV39H1) could recruit heterochromatin protein 1 homologue-γ (HP1γ) and was shown to correlate with H3K9me3 and suppression of viral transcription (Grewal SI, 2003). On the other hand, host proteins that participate in the assembling of transcriptional machinery could positively regulate viral transcription, such as nuclear factor-κB (NF-κB)(Nabel G, 1987), specificity protein 1 (SP1)(Jones KA, 1986), nuclear factor of activated T cells (NFAT)(Corthésy B, 1994) and positive transcription elongation factor b (PTEFb) (Garber ME, 1998; Zhou M, 2000). Recently, S. Goff's group identified Trim28 to be recruited to the MLV proline tRNA primer binding site (P-PBS) by adaptor protein zinc finger protein 809 (ZFP809) with suppression of MLV expression in mES via further recruiting HP-1γ to perform H3K9me2 (Teich NM, 1977; Yamauchi M, 1995; Wolf D, 2007; Wolf D, 2008; Wolf D, 2009). Upon mES differentiation, MLV expression dramatically increased as Trim28 level is positively correlated with cell pluripotency (Wolf D, 2007). G. Daley and Y. Loh's labs further showed Trim24 interacts with Trim28, which is crucial for viral silencing (data not published). This model raises the importance of Trim family proteins in silencing the retroviruses in embryonic stem cells via binding to PBS region.

I.5 Mechanism of HIV transcription restriction by viral genes interacting with host genes

Besides PBS and LTR as discussed above, some viral factors such as *Nef*(Venkatesan, 1988) and mutated *Tat*(Tyagi M, 2010) are also responsible for suppressing viral genome transcription. *Nef* was previously reported to be important for HIV pathogenesis in inducing apoptosis and negatively regulating transcription by binding to the LTR region in the SW480 cell line (Zauli G, 1999)while a *nef* mutant could maintain enhanced replication in T lymphocytes (Ahmad N, 1988). Yet these phenomena are controversial as other reports showed the opposite that deleting or inhibiting *Nef* reduces viral replication but does not reduce the cytotoxicity (Terwilliger E, 1986; Chowers MY, 1994; Spina CA, 1994; Pizzato M, 2008; Emert-Sedlak L, 2009; Ptak RG, 2010). *Nef* was also shown to facilitate the early step of viral replication by reverse transcription of the viral RNA genome (Aiken C, 1995; Chowers MY, 1995; Schwartz O, 1995). Field work showed elite controllers are infected with HIV-1 exhibiting defective *nef* alleles (Alexander L, 2000; Blankson JN, 2007; Zaunders J, 2011) and deletion of the SIV *nef* gene reduced viral replication *in vivo* and delayed onset of AIDS-like disease in macaques (Kestler HW 3rd, 1991), possibly through abolishing *Nef*'s role of down-regulating cell surface MHCI, CD4, CD8, and CD28 (Kestler et al., 1991; Jordan A, 2001; Kim YK, 2006; Blankson JN, 2007).

I.6 HIV host gene restriction in vivo--GWAS studies of European males infected with subtype-B

Host cellular factors that contribute to post-integration latency are counted for 64-83% of the total variation in virus production among individuals(Ciuffi A, 2004), indicating the importance of genome wide association studies (GWAS) to identify the host factors that control viral replication and viremia. Previous GWAS for disease progression rate showed that most hits are on HLA-B*5701 and HLA-C(Migueles SA, 2000; Fellay J, 2007; Dalmasso C, 2008; van Manen D, 2009), raising a possible mechanism of certain HLA alleles' immunity to HIV *nef* which selectively down-regulates the expression of HLV-A and –B but not –C on infected cells(Cohen GB, 1999). Other hits include ring finger protein 39 (RFN39) and zinc ribbon domain-containing 1 (ZNRD1) which encodes an RNA PolII subunit(Fellay J, 2007). Despite the above results, other meta-analysis using permutation procedures identified genes such as Notch4, C6orf12 and Trim10 (Fellay J, 2009)which were also associated with viral set point.

Chapter II A novel HIV model of human pluripotent stem cells identified Nef gene as silencing factor for provirus

II.1 Abstract

The failure of HAART on curing AIDS highlights the existence of HIV reservoir and the importance of learning host restriction to viral replication. The epigenetic restriction to HIV genome can lead to HIV latency and its mechanism remains to be studied. In this study, we used a unique model of human pluripotent stem cells to model HIV latency and reactivation, and its relationship with cell types. We found that envelope-defective full HIV genome can be silenced completely in human pluripotent stem cell. An HIV accessory protein—Nef, is able to silence the HIV in human pluripotent stem cells via interacting with host factors of NMT1, NMT2, ACOT8, Hck and PAK2. Over-expression Nef in human pluripotent cells infected with *nef*-defective HIV can silence the expressing proviral gene. The effect of Nef silencing HIV is subjective to cell pluripotency changes.

II.2 Introduction

Since 1995, highly active anti-retroviral therapy (HAART) has been widely adopted in the world. However, AIDS patients on HAART, despite maintaining low HIV viral load, cannot be cured due to HIV latent reservoirs (Gulick et al., 1997; Hammer et al., 1997; Davey et al., 1999; Imamichi et al., 2001; Palmer et al., 2008; Harrison et al., 2010). The HIV latent reservoirs are primarily in CD4+ resting T cells (CD69 and CD25 negative) (Chun TW, 1995; Chun et al., 1997; Bailey et al., 2006), yet increasing evidence indicates that latent HIV is also detected in other subsets of T cells including central memory, transitional memory, effector memory, and naïve cells (Chomont et al., 2009; Wightman et al., 2010). In addition to CD4+ T cells, there are accumulating evidence

showing macrophages (Zhang et al., 1999b; Igarashi et al., 2001), dendritic cells (McIlroy et al., 1995; Donaghy et al., 2001), monocytes (Lambotte et al., 2000; Sonza et al., 2001; Zhu et al., 2002; Ellery et al., 2007; Jaworowski et al., 2007; Hasegawa et al., 2009) and hematopoietic stem cells (Carter CC, 2010) could also serve as reservoirs for HIV replication both *in vivo* and *in vitro*.

The previous study on host restriction to HIV replications *in vitro* are fruitful yet limited in explaining how multiple epigenetic machinery is recruited onto the HIV genome that leads to *tat*-functional yet silenced HIV provirus. Other mechanisms of host silencing retroviral or retroviral elements points to the association of cell pluripotency with retrovirus silencing Murine leukemia virus (MLV) in mouse embryonic stem cells (mES) (Wolf D, 2009) and the L1 retrotransposon in human embryonic carcinoma cells (Garcia-Perez JL, 2010). In mouse ES cells,Trim28 can be recruited to the MLV Proline tRNA primer binding site (P-PBS) by adaptor protein Zinc finger protein 809 (ZFP809) and suppresses the MLV expression in mES via further recruiting HP-1γ to perform H3K9me2 (Teich NM, 1977; Yamauchi M, 1995; Wolf D, 2007; Wolf D, 2008; Wolf D, 2009). Yet in human pluripotent stem cells the mechanism of retroviral or retroviral elements latency is to be determined.

The ideal *ex vivo* HIV latency model should be the latent T cells isolated from HIV patients. However, due to horizontal transfer of activated HIV, temporal limits of *ex vivo* culture, low frequency of the latently infected cells and lack of cellular surface marker to distinguish latent cells, the mechanisms discussed above were discovered using

immortalized human cell lines infected with genetically engineered HIV. The most widely used model is Hela and Jurkat cell clones infected with a vector carrying a *tat* or *tat* H13L mutants linked to GFP(Jordan A, 2001). It performs strong HIV silencing which can be reactivated by treatment with TNF-α (Kim YK, 2006; Tyagi M, 2007). Other cell lines used include the ACH2 T-cell line (Folks TM, 1989), U1 promonocytic cell line (Folks TM, 1987), JΔK T-cell line (Antoni BA, 1994). However, transformed cells may alter the global cellular epigenetic status and the presence of viral mutations has raised questions about the relevance of using these lines to mimic latency *in vivo*. Therefore, developing better *in vitro* cell models for HIV latency is essential to probe the mechanisms of establishment of HIV reservoir.

In this study, we established an HIV latency model using human pluripotent stem cells and found that pseudotyped HIV that is defective of *nef* abolished the silencing of HIV in human pluripotent stem cells. The integrated *nef*-defective HIV can be silenced via introducing *nef* back into the cells. HIV *nef* leading to HIV latency via epigenetic modification, which is subjective to cell pluripotency status, is possibly through interacting with host factors. This study explores the molecular basis underlying the association of cell pluripotency and retroviral latency. Our result also showed a new function of *nef* in recruiting host epigenetic machinery to proviral gene. This suggests that antiviral approaches that could render *nef* gene defective should be explored as a means to reactivate latent HIV reservoir in HAART treated AIDS patients.

II.3 Material and Methods

a. Cell culture

293T cells were cultured in DMEM with 10% inactivated fetal serum (IFS), 50 U/ml penicillin, 50 mg/ml streptomycin, and 1 mM L-glutamine. All human embryonic stem cells and Induced Pluripotent Stem Cells (iPS)were maintained in undifferentiated state on inactivated primary mouse embryonic fibroblasts (MEFs) (Millipore) in hES medium (80% DMEM/F12, 20%KO Serum Replacement,10 ng/ml bFGF, 1 mM L-glutamine, 100 mM nonessential amino acids, 100 mM 2-mercaptoethanol, 50 U/ml penicillin, and50 mg/ml streptomycin). The dH1f cells were cultured in alpha-MEM containing 10% IFS, 50 U/ml penicillin, 50 mg/ml streptomycin, and 1 mM L-glutamine.

b. Virus preparation:

pNL4.3-dE-EGFP was from NIH AIDS reagent bank (catalogue no. 11100). pNL-dE-mCherry,MMLV-EGFP, MMLV-mCherry, pHR'HIV-*tat*-WT and pHR'HIV-*tat*-H13L plasmids were gift from Dr. Stephen Goff (Columbia University). 293T cells were plated at a density of 1.4 × 107 cells per 15-cm dish. The next day 293T cells were transfected with 6.25 μg viral vector, 5.625 μg lenti reprogramming viruses pSIN4-EF2-O2S and pSIN4-CMV-K2M (Addgene catalogue no. 21162, 21164), FUWdGW-Tomato (Addgene catalogue no. 22771), pHR'HIV-*tat*-WT and pHR'HIV-*tat*-H13L. For pNL-4.3-dE-EGFP and for pNL-dE-mCherry, only viral vector and VSV-G plasmid were added. Supernatant was collected 48 h and 72 h post-transfection and filtered through 45-μm pore size filters. Viral supernatants were mixed with PEG3350 solution (Sigma P3640, dissolved in PBS, 10% final concentration) for concentration and left overnight at 4 °C. Supernatants were centrifuged at 2,500 r.p.m. for 25 min, and then the pellets were

re-suspended in right amount of PBS to achieve 50x concentrated viruses. All virus titration was performed on 293T cells.

c. SYBR-Green real-time RT–PCR

Total genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen). Total RNA was extracted using TRIzol and RNeasy Mini kit coupled with an RNase-free DNase set (Qiagen) and reverse transcribed by Superscript III (Invitrogen). The resulting complementary DNAs were used for PCR using SYBR-Green Master PCR mix (Applied Biosystems) in triplicates. All quantifications were normalized to an endogenous β-actin control. The relative quantification value for each target gene compared to the calibrator for that target is expressed as $2^{-(Ct - Cc)}$ (Ct and Cc are the mean threshold cycle differences after normalizing to β-actin).The H9 HIV was amplified by primers annealing to LTR region. For iPS colonies, the HIV was amplified by primer annealing to the gag gene. Single copy control of HIV-Δ*env*/*nef* is 8E-5 cell line, which is a subclone of HIV-infected CD4+ CEM-derived human T-cell line with single copy of viral gene insertion and confirmed by southern blot. Single copy control of HIV-SIN was obtained by single cell clone of 293T infected with MOI=0.1 viruses and 80% of the colonies reached the lowest level of viral integration by gDNA qPCR (data not shown), indicating that they are single copy according to Poisson distribution.

d. Reprogramming assays

The reprogramming assay was performed as described (Onder TT, 2012) with modification: dH1f cells were first infected with MLV viruses at high multiplicity for

labeling and HIV viruses at low multiplicity of infection (m.o.i=0.1) to ensure majority of infected cells did not receive= more than one HIV vector. Twenty-five thousands infected dH1f cells were then plated per well in 12-well plates and infected overnight with lentiviral (100–200 µl supernatant before concentration)(Yu et al., 2007). For late infection group, the infected dH1f with MLV were re-plated at density of 25,000 cells per well in an 12 well plates and infected with low MOI of HIV five days post- infection with reprogramming viruses. Six days post- infection with OSKM lentiviruses, cells were trypsinized and re-plated 1:6 onto six-well plates with MEF. Medium was changed to hES medium daily until day 28 when colonies were picked. All colonies were split for Tra-1-60 staining before expanding.

e. Immunostaining

The iPS colonies cells were fixed with 4% paraformaldehyde and stained with biotin-anti-Tra-1-60 (eBioscience, catalogue no. 13-8863-82, 1:250) and streptavidin horseradish peroxidase (Biolegend, catalogue no. 405210, 1:500) diluted in PBS (3%), FCS (0.3%) and Triton X-100. Staining was developed with the Vector labs DAB kit (catalogue no. SK-4100).

f. Integration site sequencing

Genomic DNA (2.5ug) were fragmented using DraI in 100ul reaction at 37℃ overnight, followed by pheno-chloroform extraction. Digested DNA was ligated to an adaptor in Universal Genome Walker kit (BD Biosciences Clontech, Palo Alto, CA). The ligation was done in a 8ul reaction at 16℃ overnight. First and second round PCR were

performed using primers specific to HIV 3'LTR (Ciuffi and Barr, 2011) and AP1/AP2 provided by Universal Genome Walker kit(BD Biosciences Clontech, Palo Alto, CA). Both PCR were performed in 25ul reaction and 1ul of PCR1 products were used for second round PCR. The PCR amplicons were purified using QIAquick PCR Purification Kit (Qiagen) before used for Illumina MiSeq. The BLAT program (University of California, Santa Cruz, Human Genome Project working draft Feb 2009; http://www.genome.ucsc.edu/) (Karolchik et al., 2009) was used to map integration sites. We include only amplicons with the three criteria (MacNeil et al., 2006) into the analysis: (i) contained the terminal 3'end of the HIV-1 LTR; (ii) matched genomic DNA within five bps of the end of the viral 5' LTR; (iii) matched a single human genetic locus with at least 95% homology across the entire sequenced region. The genes that HIV is integrated into (exon, intron) were further analyzed for their expression level in human pluripotent stem cells.

g. Nef cloning

The pSIN-EF2-Lin28A-puro (Addgene, catalogue no. 16580) is digested with EcoRI and BamHI for 3 hours at 37℃ and purified using QIAquick Gel Extraction Kit (Qiagen). HIV-1 *nef* gene and EGFP gene were amplified from pNL4-3-dE-EGFP plasmid with primers carrying EcoRI and BamHI sites. Amplicons were subcloned into pCR2.1 using TOPO® TA Cloning® Kit for Subcloning (Invitrogen, 450641) and digested with EcoRI and BamHI. After gel purification, inserts and vectors were ligated at room temperature for 1 hour using T4 DNA Ligase (1 U/μL) (Invitrogen, 15224017). The ligated plasmids

were then amplified using One Shot® Stbl3™ Chemically Competent E. coli (Invitrogen, C737303) .

h. Western blot

The Western blot was performed as described (Zhu et al., 2011). Primary antibodies used were anti-Nef (Millipore, MAB899) and anti-alpha tubulin (Cell Signaling, #3873) antibodies. Secondary antibodies were sheep-anti-mouse or donkey-anti-rabbit HRP-conjugated secondary antibody (GE Healthcare).

i. ChIP-PCR:

ChIP analyses were performed as described (Cesana et al., 2011). Briefly, chromatin was extracted from cell culture according to manufacturer's specifications (MAGnifyChIP - Invitrogen) with the following antibodies: Anti-trimethyl-Histone H3 (Lys27) antibody (Millipore, 07-449), Anti-acetyl-Histone H4 antibody (Millipore, 06-598), Anti-trimethyl-Histone H3 (Lys4) antibody (Millipore, 07-473), Anti-trimethyl-Histone H3 (Lys9) antibody (abcam, ab8898). A standard curve was generated for each primer pair testing 4-point dilutions of input sample. Fold enrichment was quantified using qRT-PCR(QuantiTect SYBR Green - QIAGEN) and calculated as a percentage of Inputchromatin (% Inp).

j. EB differentiation:

The EB differentiation is performed as described(Cerdan et al., 2007). Basically the iPS cells infected with HIV were collected as large aggregates and resuspended in EB

differentiation medium composed of 80% DMEM, 20% FCS (Stem Cell Technologies 06900), 50 Pg/ml ascorbic acid, 0.2Pg/ml holotransferrinon low attachment dishes. The next day, cytokines were added: hSCF (300 ng/ml), hFlt3L (300 ng/ml), IL-3 (10 ng/ml), IL-6 (10 ng/ml), G-CSF (50 ng/ml), BMP4 (50 ng/ml). Media containing cytokines was replaced every three days for 14-16 days. Then the EB were disassociated as described(Cerdan et al., 2007).

k. Myeloid and Erythroid lineage differentiation

The disassociated EB was sorted as described(Doulatov et al., 2013). Sorted cells were resuspended in IMDM with 2% IFS at a concentration of 10K/300ul. Five thousand cells were plated with 1.5ml MethoCult™ (StemCell Technologies H4434) and CFU assay was performed according to manufacturer's protocols. After 14 days the cells were harvested and flow cytometry was performed as described(Doulatov et al., 2013).

l. Fibroblast differentiation

Pluripotent H9 or iPS cells were passaged onto plates coated with gelatin (0.2%) by CollagenaseIV. Cells were cultured in medium containing 80% KO DMEM (Invitrogen 10829-018), 20% FCS, 0.67% 2-ME, 50ug/ml Ascorbic acid, 200ug/ml h-transferrin (Sigma T0665-1G). The cells were passaged for four passages with trypsin when reaching confluence. The differentiated cell lines were confirmed by both morphology and the loss of the pluripotent marker (Tra-1-60).

II.4 Result

a. HIV-Δ*env* is silenced at transcriptional level in human pluripotent stem cells

Murine leukemia virus (MLV) undergoes silencing in a time dependent manner in mouse embryonic stem cells(Schlesinger and Goff, 2013), and can also be well-silenced in human pluripotent stem cells (Kamata et al., 2010). Whether full-length HIV can also be silenced in embryonic stem cells is unknown. To study the extent the HIV is silenced in human pluripotent stem cells, we used an envelope-interrupted full-length HIV-1 vector(Zhang et al., 2004) (Figure 5A), pseudotyped with VSV-G envelope to infect human embryonic H9 cells (WiCell, Madison, WI).

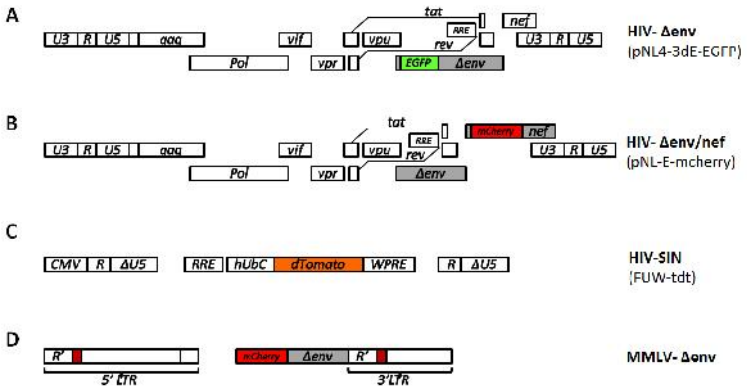

Figure 5. Retroviral constructs used in the experiments.
Original plasmid name was indicated in parentheses. (A) A full-length HIV backbone with envelope interrupted with EGFP(Zhang et al., 2004). (B) A full-length HIV backbone with a frame-shift mutation on envelope and interrupted Nef gene with mCherry. (C) A self-inactivating lentiviral vector with hUbC promoter driven dTomato. (D) A Moloney murine leukemia virus LTR driven mCherry vector, depleted with gag/pol and with an inactivated envelope.

To better observe the MOI threshold of viral silencing, the infection was done with MOI starting from 0.005 to 2.5 (titrated by 293T, Figure 6A) and percent of HIV positive cells expressing HIV were measured by FACS on day 12 post infection. Cell counts were also

surveyed to monitor cytotoxicity. Due to the limitation of FACS to differentiate the cytotoxic effect versus viral silencing effect, which will both show reduced percent positive cells, the viral DNA were collected on day 12 post infection and the integration number were measured by qPCR (Figure 6A). As a positive control, we used an MLV reporter vector (Figure 5D) that is known to be silenced in human pluripotent stem cells. As a negative control we used an self-inactivated lentiviral vector (HIV-SIN) with a ubiquitinated hUbC promoter that is known not be silenced in human pluripotent stem cells (Figure 5C)(Ma et al., 2003; Rompani and Cepko, 2008; Norrman et al., 2010). We found that compared to controls, HIV-Δenv is completely silenced in H9 cells at all MOI day 12 post infection (Figure 6B) with a comparable level to MLV. While all viruses are cytotoxic at high MOIs, HIV-Δenv is cytotoxic at MOI below 0.1 both upon infection and over time(Figure 6D, Figure 6G).

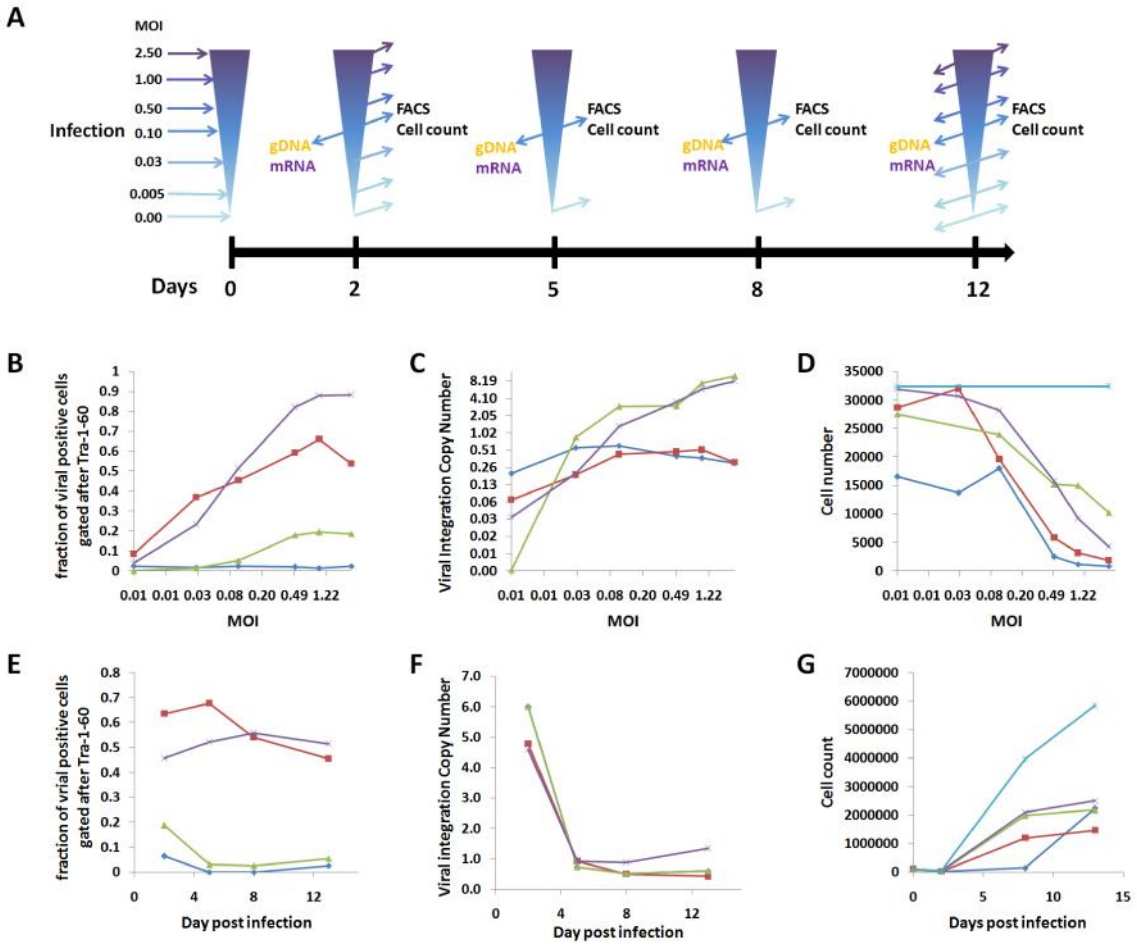

Figure 6. HIV-Δ*env* is transcriptionally silenced and cytotoxic to human pluripotent stem cells. (A) Schematic presentation of viral infection into H9 cells. (B) Percent of positive cells gated on Tra-1-60 positive cells day 12 post infection. Dark blue: HIV-Δenv, red: HIV-Δ*env/nef*, green: MMLV, purple: HIV-SIN. (C) The viral integration level day 12 post infection. Single copy control of HIV-SIN was obtained by single cell clone of 293T infected with MOI=0.1 viruses and 80% of the colonies reached the lowest level of viral integration by gDNA qPCR (data not shown), indicating they are single copy according to Poisson distribution. Single copy control of HIV-Δ*env/nef* is 8E-5 cell line, which is a subclone of HIV-infected CD4+ CEM-derived human T-cell line with single copy of viral gene insertion and confirmed by southern blot. Color representatives are the same as A. (D) Cell count of day 2 post infection, X-axis is on a log2.5. Both HIV-Δ*env/nef* and HIV-Δenv showed cytotoxicity at high MOI yet only HIV-Δenv showed significant silencing at low MOI. (E) Percent of positive cells gated on Tra-1-60 positive cells infected at MOI=0.1. Color representatives are the same as A. (F) The viral integration level at MOI=0.1 infection surveyed on different days post infection. Color same as A. (G) cell count of infection at MOI=0.1; Light blue: the Mock infection. Other color representatives same as A. D and G showed that though all viruses are killing cells at high MOI, HIV-Δenv is extremely toxic to cells even at low MOI. The x-axis is shown on a log2.5 scale to reveal low MOI. Viral integration level in H9 cells negative fraction sorted on Day 12 post infection by MOI=0.1 viruses.

b. Evaluating the silencing kinetics of envelope interrupted Full-length HIV(HIV-Δ*env*) in human pluripotent stem cells

To better dissect the silencing phenotype, the percent positive cells and proviral integration level were also surveyed on different days (Day 2, 5, 8 and12) post infection for the infection of MOI=0.1 (Figure 6A). For all vectors, the proviral integration level decreases dramatically from an average high copy number on day 2 to copy numbers as low as one on day 5 (Figure 6F, Figure 7B), indicating cytotoxicity of all vectors at high MOI and the fact that H9 is more infectable than 293T. The percent of cells expressing HIV-Δenv decreased over time (Figure 6E) and correlated with the decrease of integration level over time (Figure 6F), providing little evidence of overtime silencing of HIV-Δenv vector (Figure 6E, Figure 6F, Figure 7B). Therefore the decrease in positive cells of HIV-Δenv derived vectors is likely due to cytotoxic effect.

To further confirm that the observed post-integration viral silencing is at transcriptional level rather than post transcription level, the mRNA of cells infected with MOI=0.1 were collected at different days post infection (Figure 6A) and viral transcript level was measured by quantitative reverse transcription PCR (qRT-PCR). The qRT-PCR results measuring the HIV-Δenv viral mRNA expression level at different day post infection and different MOI all correlate well with the percent positive cells (Figure 7B), indicating all silencing is exerted at transcription level.

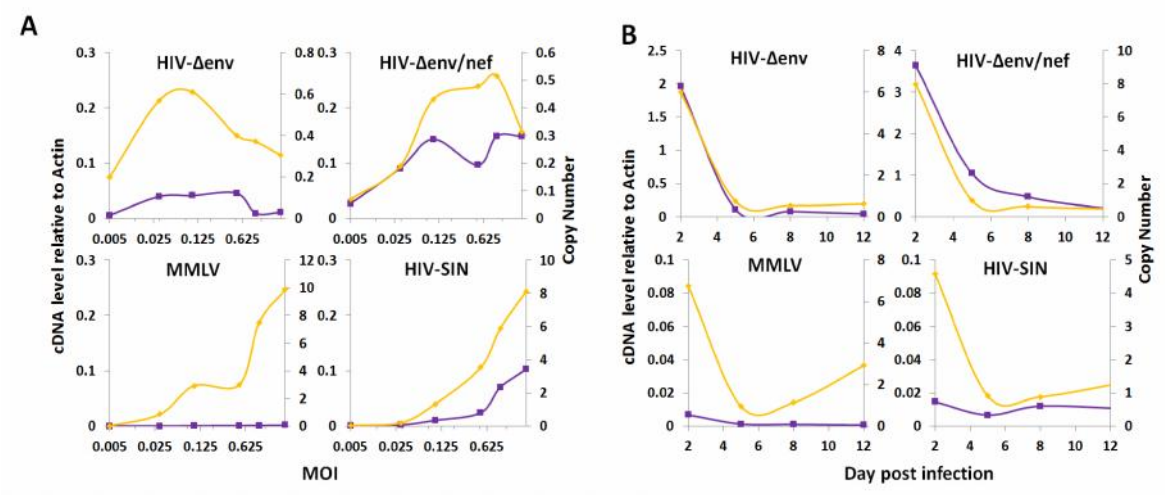

Figure 7. HIV-Δ*env* is transcriptionally silenced in human pluripotent stem cells while HIV-Δ*env*/*nef* is not.
(A) Insertion and expression of viral genes extracted from H9 cells infected with different virus of increasing MOI, day 12 sampling. Orange: genomic DNA shown in copy number; Purple: cDNA shown relative to Actin. This figure showed that compared to HIV-SIN (which is not silenced) and MMLV(which is silenced), HIV-Δ*env*/*nef* shows no sign of silencing while HIV-Δenv shows some degree of silencing in H9. (B) Insertion and expression of MLV and HIV based vector proviral genes from H9 infected with MOI=0.1 viruses and sampled on different days post infection. Orange: genomic DNA shown in copy number; Purple: cDNA shown relative to Actin. This figure presents that cDNA level of MLV correlates with the positive percent of cells while gDNA doesn't, meaning that over time the reduction in positive percent of cells is largely due to silencing of the virus. Compared to MLV, the level of HIV-Δenv cDNA correlates with gDNA and the percent of positive cells, meaning that there is no overtime silencing effect and the reduction in positive percent cells is all due to the cytotoxicity.

c. Knocking out different HIV accessory genes affect the extent of viral silencing in human pluripotent stem cells.

To study the role of HIV accessory proteins in silencing HIV in human pluripotent stem cells, we infected the H9 cells separately at MOI=0.1 with the four vectors shown in Figure 5 and two vectors shown in Figure 8A (Pearson R, 2008). We chose these vectors because the vector in Figure 5B and the two vectors in Figure 8A have multiple accessory protein deleted including *nef*, *vif*, *vpr*. After culturing for 12 days, we collected the cells and sort for Tra-1-60+ and viral fluorescence negative fraction. Genomic DNA is collected for the negative fraction. We found that infected with similar amount of viruses,

there are less proviral genes in the negative fraction, indicating that deletion of vif, *vpr* and *nef* results in partially active HIV expression (Figure 8). Significantly, the HIV vector which has an interrupted *nef* gene via mCherry and inactivated *env* otherwise same with HIV-Δenv (Figure 5B) is completely depleted of viral integration in the negative fraction, meaning that there is no residue HIV-Δ*env*/*nef* in the negative fraction of MOI=0.1 infection (Figure 8).

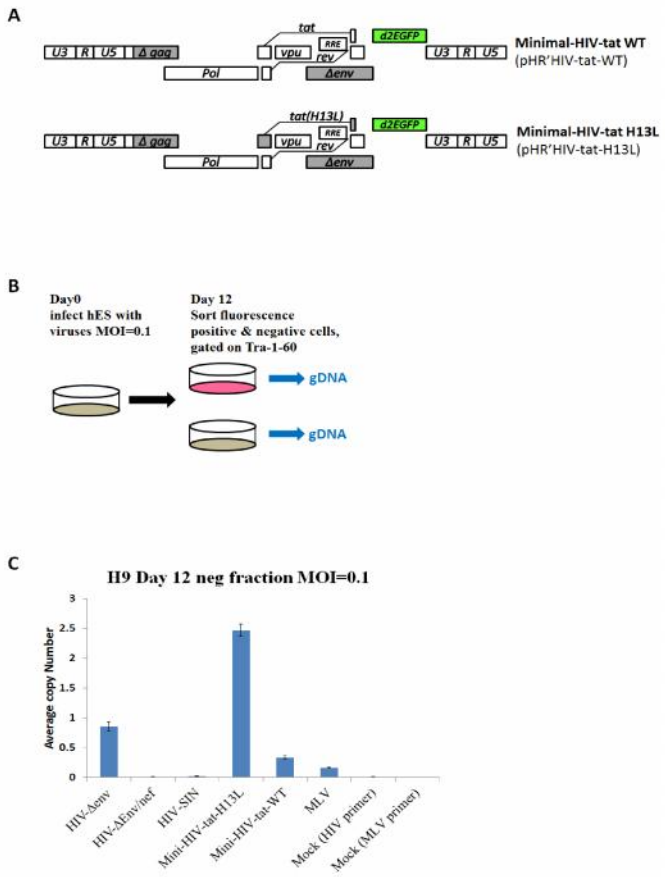

Figure 8. Deleting HIV accessory proteins reactivates HIV by cell sorting.
A. The constructs with HIV accessory protein deletions. Upper: HIV construct without *vif*, *nef*, *vpr* but has a functional *tat*. Lower: Tat is inactivate with a H13L mutation otherwise same as upper. B. Schematic presentation of cell sorting strategy. C. Genome DNA qPCR of negative fraction sorted on Day12 post viral infection. Copy number was calculated based on qPCR result of single copy 8E-5 cells. As expected, mutating that would cause viral silencing, while deleting multiple HIV accessory genes result in partially proviral gene reactivation. Significantly, HIV-Δ*env*/*nef* vector has no integrated virus in the negative fraction.

d. Interrupting Nef at genomic level abolishes the transcriptional silencing and cytotoxicity of HIV

To better document expression pattern of HIV-Δ*env*/*nef* in human pluripotent stem cells, we performed the experiment shown in Figure 6A with HIV-Δ*env*/*nef*. Strikingly, compared to HIV-Δenv, HIV-Δ*env*/*nef* has similar proviral integration level (Figure 6C), yet is highly expressed in H9 cells even at MOI lower than 0.1 and over time (Figure 6B, Figure 6E). The percent of cells expressing HIV-Δ*env*/*nef* is at a comparable level to HIV-SIN (Figure 6B, Figure 6E), indicating the lack of transcriptional restriction in H9 cells to HIV-Δ*env*/*nef*. While the viral expression is compromised at high MOI by comparing HIV cDNA and gDNA (Figure 7A), RT-qPCR result showed a high correlation between the mRNA and gDNA level of HIV-Δ*env*/*nef* over days at MOI=0.1 (Figure 7B). Our data suggest that deletion of *nef* gene from HIV genome can completely reverse transcriptional silencing of provirus in human pluripotent stem cells. Additionally, deleting *nef* abolished the cytotoxicity of HIV-Δenv at low MOI and over time. Interestingly, we also observed that both HIV-Δ*env* and HIV-Δ*env*/*nef* showed a saturation of integration level compared to HIV-SIN and MLV, indicating strong pre-integration restriction at high MOI to full-length HIV genome.

e. Establishing single integrated HIV pluripotent stem cells via somatic reprogramming

MLV is rapidly silenced upon somatic reprogramming (Kamata et al., 2010). To test whether reprogramming fibroblast infected with HIV-Δenv can also recapitulate the phenotype of silencing HIV genome, and to establish a colonic pluripotent stem cell model with single copy of HIV integrated, we performed reprogramming experiment. We reprogrammed HIV infected dH1f cells and generated iPS cells comparable to H9 cells in regard to cell pluripotency(Park IH, 2008). We infected human differentiated fibroblast cells dH1f (Park IH, 2008) with low MOI of MLV and the HIV vectors (Figure 5) carrying different fluorescence color, and sequentially infect the cells with high MOI of Oct4, Sox2, c-Myc and Klf4 delivered by HIV-SIN vectors(Yu et al., 2009) (Figure 9A). Before reprogramming, the dH1f cells showed positivity for both HIV and MLV. Four types (MLV positive or negative, HIV positive or negative) of colonies emerged upon day 28 of reprogramming. MLV is completely silenced in all the colonies stained positive for Tra-1-60 (Figure 9B, data not shown) which proved that MLV is well silenced by reprogramming dH1f, while HIV-SIN-hUbC is highly expressed in the colonies (Figure 9B). For HIV-Δenv and HIV-Δ*env*/*nef* reprogramming experiments, we randomly picked 50 colonies from each and screened for HIV viral positivity via genomic DNA qPCR and obtained colonies that are inserted with either HIV-Δenv or HIV-Δ*env*/*nef* alone(Figure 9D). We found that all the HIV-Δ*env*/*nef* colonies highly expressed the HIV gene except for a few that has *tat*-mutation in *tat*-exon1 (generating a premature stop codon in the middle of exon 1) bearing a silenced HIV-Δ*env*/*nef*, while all colonies silenced the MLV. However HIV-Δenv is completely silenced in the integrated colonies. Colonies that has a single infection of HIV-Δ*env*/*nef* have similar HIV expression phenotypes (data not shown).We further sequenced the proviral gene and confirmed that except for the *tat*

mutated colony (iPS C3),the other integrated HIV proviruses in these colonies are intact without mutations (data not shown). Our result showed that HIV-Δenv can be silenced via somatic reprogramming while HIV-Δ*env/nef* cannot.

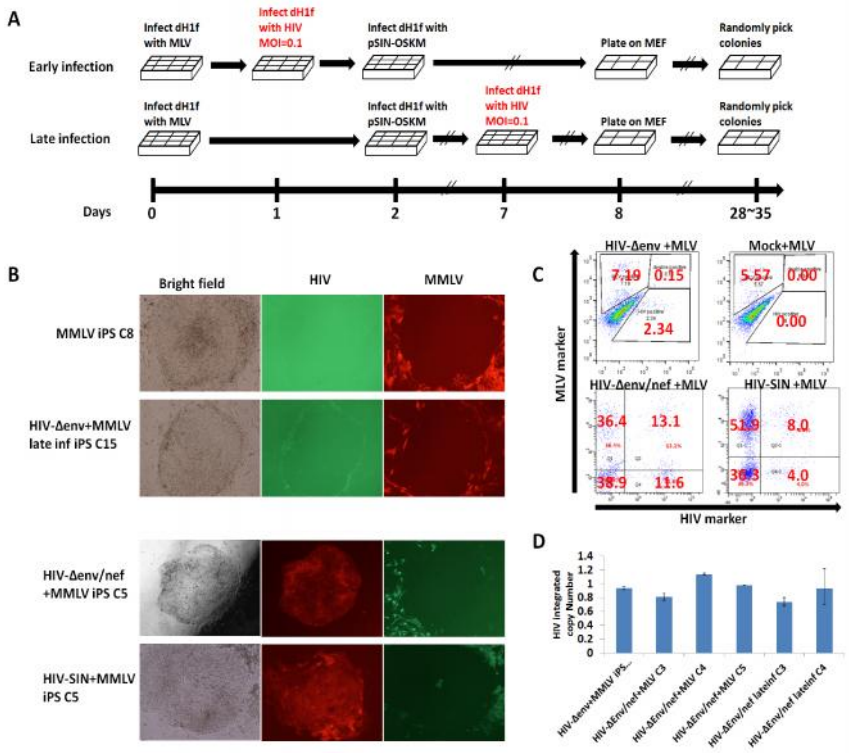

Figure 9. HIV-Δenv is silenced upon somatic reprogramming while HIV-Δ*env/nef* is not
A. Experiment design of reprogramming. B.Tra-1-60 stained positive colonies infected with four vectors. Both the HIV-Δenv and MLV are well silenced upon reprogramming, yet HIV-Δ*env/nef* and HIV-SIN-hUbC are not. C. FACS of the doubly infected dH1f cells before reprogramming. dH1f cells were infected with MLV and HIV before reprogramming. These cells were subjected to infection with SIN-vector encoding OSKM genes later. D. Viral integrated iPS colonies. To confirm the presence of viral genes in the iPS colonies containing HIV gene, qPCR of HIV genes were performed.

f. ChIP-PCR revealed the epigenetic signature of silenced HIV proviral genome in pluripotent stem cells.

To investigate the mechanism of HIV proviral gene silencing in human pluripotent stem cells, we collected the chormatins of iPS cells with a single integration of silenced HIV-Δenv (iPS C15, Figure 9D) or actively expressing HIV- Δ*env/nef* (iPS lateinfC3, Figure

9D) and performed ChIP-PCR assay to investigate the epigenetic features on HIV proviral gene. We used primers annealing to LTR region of HIV proviral genes, as well as *env* region of HIV proviral genome to detect the histone marker enrichment on HIV genome. Two human genes, GAPDH and Pax3, were used as control. We found that compared to HIV-Δenv, HIV-Δ*env*/*nef* is approximately two-fold more enriched for H3K4me3 (Figure 10B), which are associated with active gene expression in human pluripotent stem cells (Pan et al., 2007; Garcia-Perez JL, 2010; Hawkins et al., 2010). H3K9me3 is 1.5 fold more enriched on HIV- Δenv compared to HIV-Δ*env*/*nef*, while H3K27me3 and H4Ac are similarly enriched on HIV genome in either HIV-Δ*env*/*nef* or HIV-Δenv colonies (Figure 10B). We obtained similar result using primers annealing to HIV *env* region on proviral gene (Figure 10A) indicating the histone modification is presence throughout HIV genome rather than solely on LTR. We conclude that HIV-Δenv is silenced in human pluripotent stem cells via enriched H3K9me3, and lack of H3K4me3.

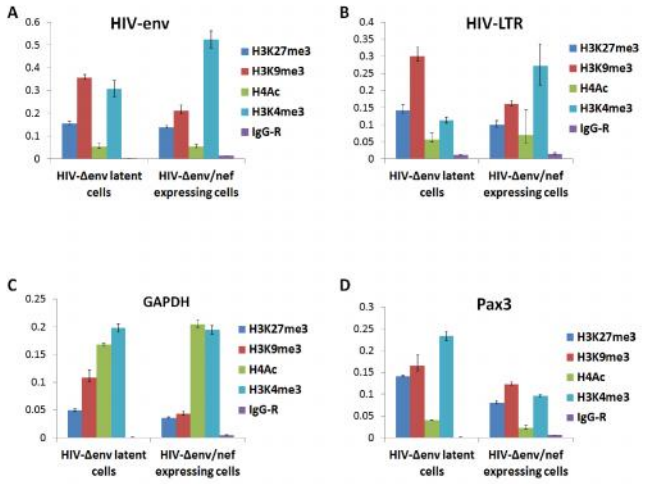

Figure 10. Epigenetic signature difference between HIV-Δenv and HIV-Δ*env*/*nef* proviral LTR in human pluripotent stem cells.
ChIP-based measurement of H3K9me3, H4Ac, H3K27me3 and H3K4me3 on (A) HIV *env* gene, (B) HIV LTR, (C) GAPDH, (D) Pax3. The signal was normalized to Input (Y axis shows the percentage of signal relative to input). GAPDH gene should be enriched for H3K9me3, H3K4me3 and H4Ac, but not H3K27me3(Hawkins et al., 2010). Pax3 should be enriched for H3K27me3 and H3K9me3 but limited

enrichment for H4Ac(Hawkins et al., 2010).

g. Nef suppressing HIV expression is not through genome-wide integration sites selection

Our data suggested that HIV-Δenv is silenced in human pluripotent stem cells epigenetically. However, due to the fact that HIV-Δenv is more cytotoxic than HIV-Δ*env*/*nef*, it is also possible that Nef induce HIV silencing indirectly via selection for proviral integration sites that have less permissive chromatin structures. To investigate whether Nef regulates HIV expression via chromatin modification indirectly by inducing cell death, we sequenced the integration sites of HIV-Δenv and HIV-Δ*env*/*nef* MOI=0.1 infection on day 2, 5 and 12 post infection (Figure 6A) using MiSeq (Illumina) deep sequencing technology. The strategy of recovering proviral integration sites and pooling samples for deep-sequencing is showing in Figure 11.

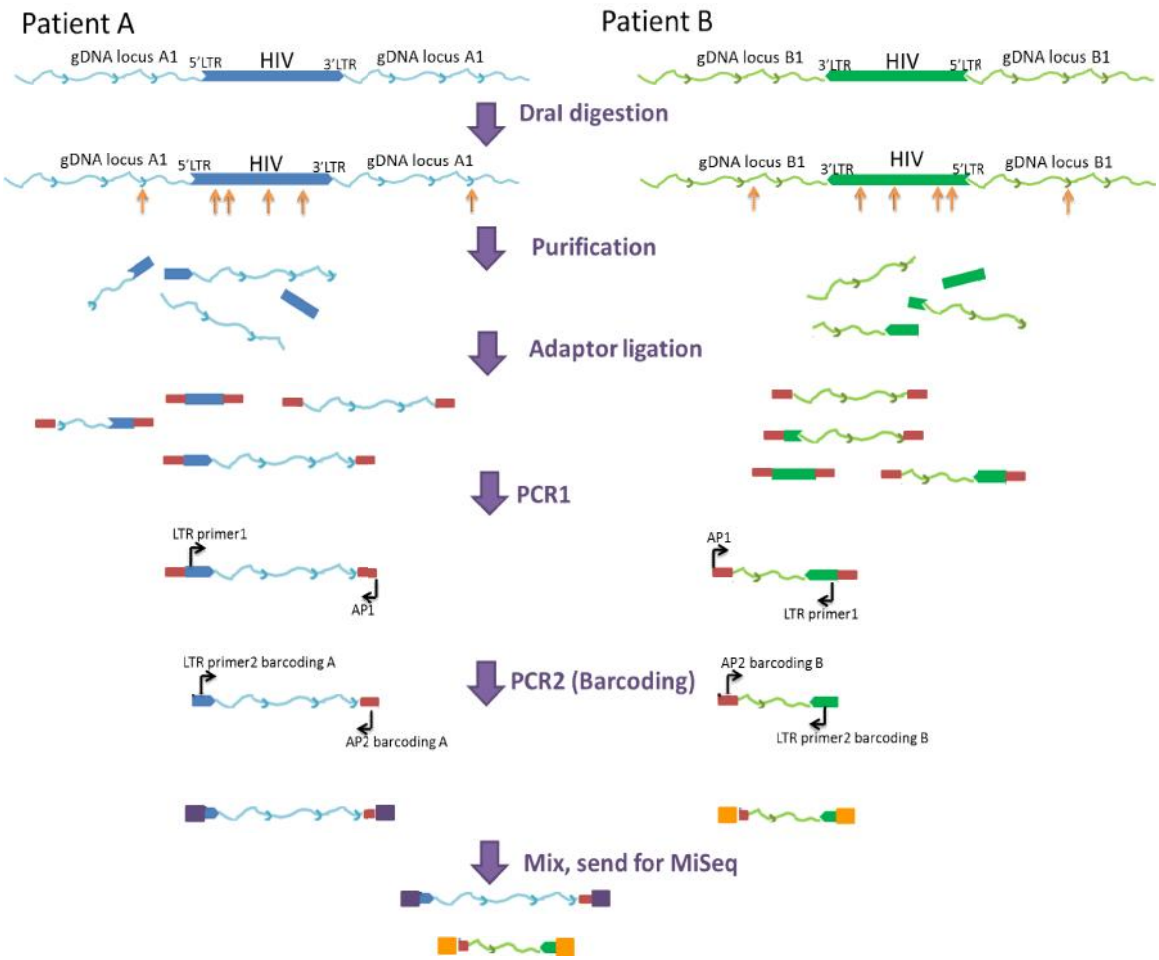

Figure 11. High-throughput integration sites sequencing strategy.
The genomic DNA were digested using DraI enzyme overnight. The fragments were purified and ligated with an adaptor. Two rounds of PCR were conducted using one primer that anneal to the adaptor and the other one primer annealing to viral LTR sequence. The second round of PCR used primers that have a 5-bp barcode on it. The samples were then mixed and sent for MiSeq.

We found 95% of the provirus are integrated into genes, and HIV is more prone to integrate into highly expressed genes compared to background (Figure 12A). Compared to HIV, the distribution of expression level of the genes integrated by MLV is very similar to background distribution. Comparing HIV-Δenv with HIV-Δ*env*/*nef*, there was no significant difference in the orientation of the proviruses inserted into host genes. Across all the HIV vectors, we found the genes integrated by HIV have similar expression distribution (Figure 12A). Therefore, we conclude that Nef suppresses HIV

proviral gene expression not through cytotoxic effects induced by integration site selection.

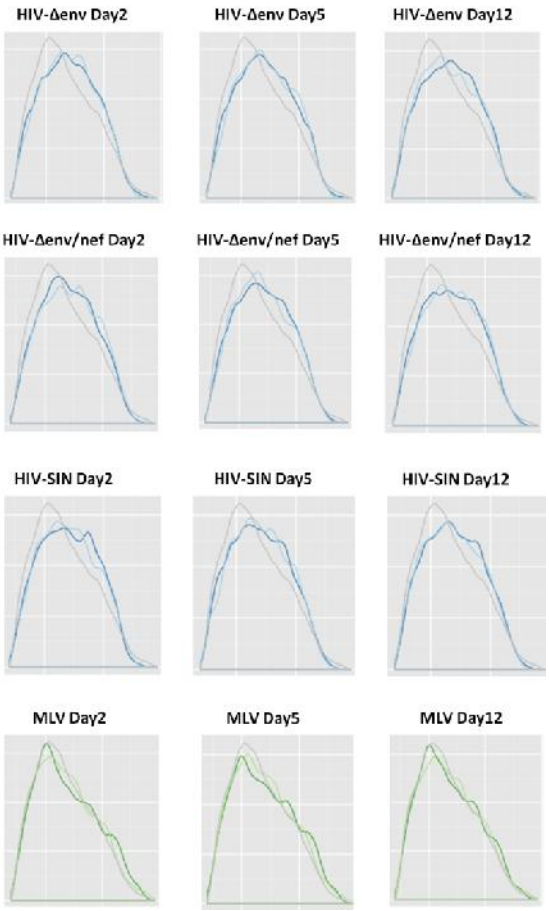

Figure 12. HIV-Δenv and HIV-Δ*env/nef* are integrated into similar regions of human genome.
The histogram of the expression score of genes that contain HIV provirus. Score is based on microarray data of H9 cells (GSE45036).Dark blue/green: the HIV/MLV genome orientation is same as the host genes. Light blue/green: the HIV/MLV genome orientation is opposite to the host genes.

h. Over-expressing Nef imposes silencing to HIV proviral gene in human pluripotent stem cells

To investigate the role of Nef in silencing HIV proviral gene and to confirm that Nef along is able to contribute the phenotype, we made the Nef expression construct (collaterally with an EGFP expressing construct) from HIV-SIN based lenti-vector (Ma et al., 2003; Yu et al., 2007) with an internal promoter of EF1α (Figure 13B) that can be

expressed in human pluripotent stem cells (Norrman et al., 2010)in trans. We performed the Nef supplementing experiment shown in Figure 13A and confirmed Nef expression by qPCR (data not shown). Basically, we co-infect H9 with a Nef donor HIV based vector (pSIN-Nef, HIV-Δenv, or negative control of pSIN-EGFP), together with HIV-Δ*env*/*nef*. By co-infecting the H9 with pSIN-Nef and HIV-Δ*env*/*nef*, and surveying the HIV expression via FACS on Day 5 post infection, we observed reduced HIV expression by three-fold compared to infected with EGFP (Figure 13C). Co-infecting HIV-Δ*env*/*nef* with HIV-Δenv can also silence the HIV-Δ*env*/*nef* (Figure 13C), indicating Nef can silencing the proviral genome in trans.

Since we observed less lenti-viral expression efficiency at high MOI infection with multiple copies in one cell (Figure 6 and Figure 7), the copy numbers of both the HIV-Δ*env*/*nef* and the Nef donor lenti-vector can potentially affect the expression level of the viruses. Therefore, we measured the copy number of HIV-Δ*env*/*nef* and the Nef-donor virus to make sure the cells have a similar level of HIV integration (Figure 13E) as well as Nef-donor viruses.

By infecting the iPS colonies that stably expressing HIV (iPS lateinfC4, Figure 9D) with pSIN-Nef and surveying the HIV expression via FACS on Day 12 post infection, we observed a 6-fold decrease in viral expression (Figure 13D). Our data strongly suggest that over-expressing Nef is able to silence the actively expressing HIV-Δ*env*/*nef* in human pluripotent stem cells.

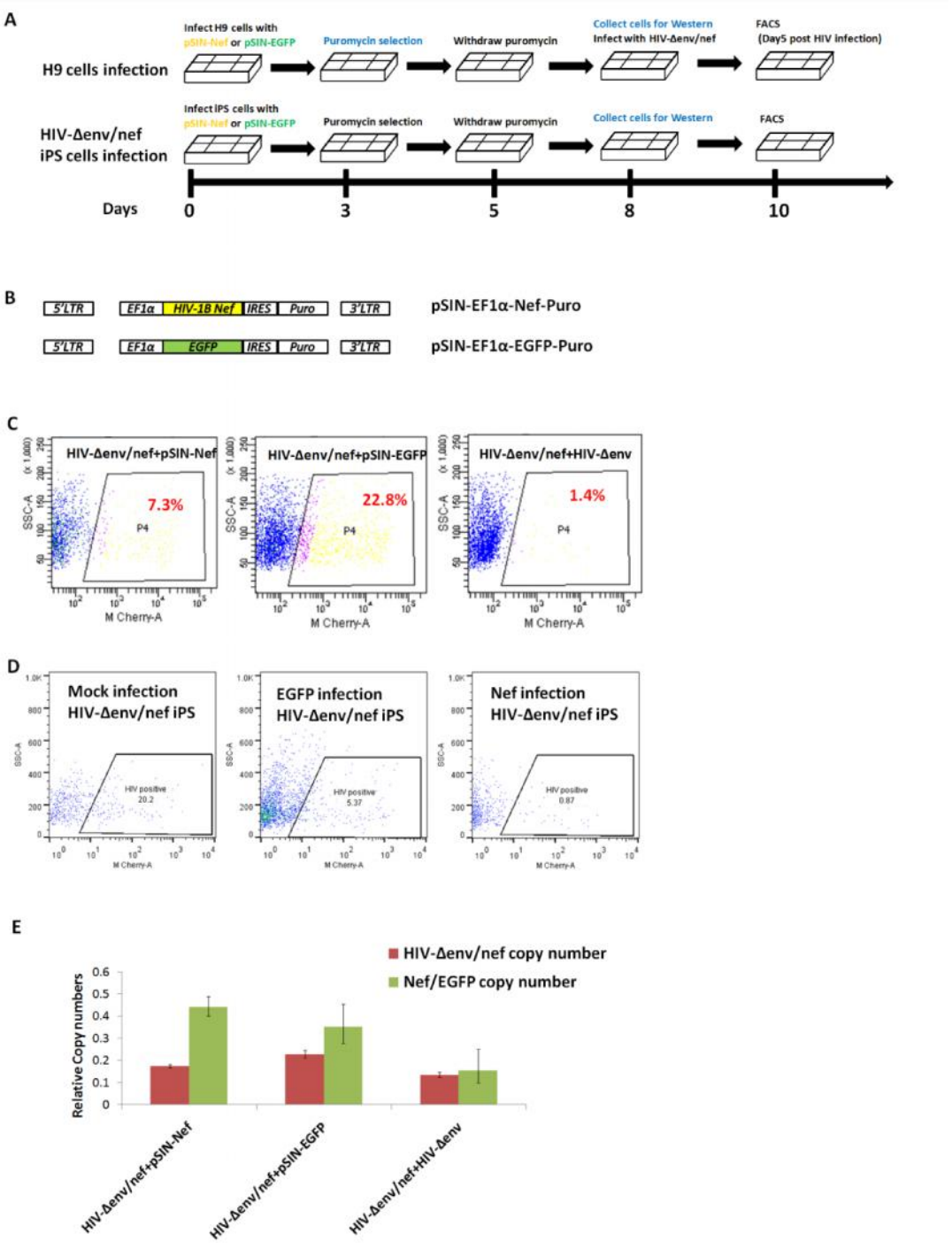

Figure 13. Supplementing *nef* into HIV-Δ*env*/*nef* infection can silence the virus in human pluripotent stem cells.

A. Schematic presentation of experiment design. B. Nef and EGFP construct made from pSIN-EF1α-Lin28A-Puro (3). C. FACS result showing fraction of positive cells in pSIN-Nef and HIV-Δ*env*/*nef* co-infection in H9 cells, Day5 post infection. D. FACS result showing fraction of positive cells in pSIN-Nef infected iPS cells bearing single HIV-Δ*env*/*nef* integration (iPS C4, Figure 7D).E.Copy numbers of HIV-Δ*env*/*nef* and the lenti-Nef in H9 infection.

i. HIV expression and Nef-induced proviral silencing is linked to cell pluripotency

To further answer the question that whether Nef-linked HIV latency occurs only in pluripotent cells, we infected several cell types (H9, 293T, dH1f and Jurkat) with MOI=1 HIV-Δenv and performed cell sorting on Day2 post infection. We surveyed the genomic DNA integration level of the cells in that doesn't express fluorescence proteins integrated in HIV genome (FACS negative fraction). HIV-Δenv gene is present in all cells studied(Figure 14A), meaning that the HIV-Δenv latency is universal in all four cell types surveyed. However, comparing the amount of viruses in positive fraction (represented by percent positive cells) and viruses in negative fraction (represented by copies of integrated viruses in negative fraction), HIV-Δenv is silenced to different extent in cells with different pluripotency, especially H9 cells showed the strongest degree of silencing the virus while the Jurkat cells showed the least (Figure 14A). To better understand the relationship between cell pluripotency and viral silencing, we induced differentiation of iPS cells inserted with single copy of HIV-Δenv (iPS C15, Figure 9D) into fibroblast. We found a 6.7% of viral reactivation after inducing differentiation to fibroblast cells (Figure 14C). Our results indicate that human pluripotent stem cells have a stronger ability to maintain the HIV-Δenv proviral gene in silence compared to dH1f, 293T and Jurkat.

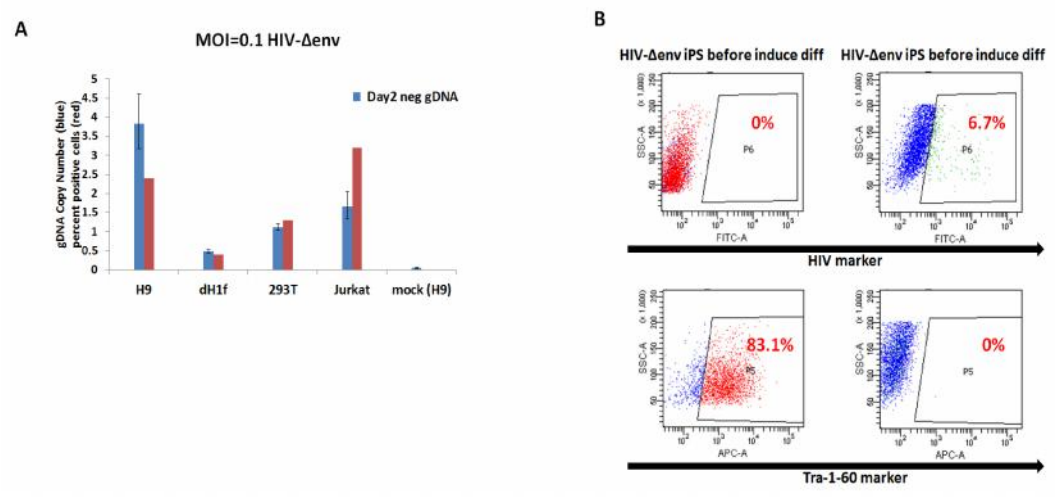

Figure 14. HIV-Δenv latency is linked to cell pluripotency.
A. Cell sorting experiment of HIV-Δenv infecting cells with different pluripotency. H9, dH1f, 293T and Jurkat cells are infected with HIV-Δenv at MOI=1 (titrated by 293T), and sorted for negative fraction on Day2 post infection. Genomic DNA is collected from the cells in the negative fraction. HIV integration copy number was obtained by qPCR compared to single copy 8E-5 cells. Blue bar: genomic DNA integration level in the negative fraction, representing amount of viruses in negative fraction. Red bar: percent of positive cells, representing amount of viruses in positive fraction. B. FACS result of inducing differentiation to fibroblast cells from HIV-Δenv integrated iPS cells.

As macrophages and multiple cells in blood lineages are nature host of HIV, and since Jurkat cells has least ability to silence HIV-Δenv, we asked whether inducing differentiation to blood lineage can also reactivate the silenced HIV proviral gene. We induced differentiation of HIV-Δenv integrated iPS colony (iPS C15, Figure 9D) to blood lineage by a two step process. We first successfully obtained CD34+CD45+ hematopoietic stem cells or CD34+ endothelial progenitor cells from iPS cells via embryoid body formation assay. The silenced HIV-Δenv is not reactivated in both population(data not shown). We then performed colony formation assay to both CD34+ and CD34+CD45 population and obtained terminally differentiated blood cells including granulocytes, erythrocytes and macrophages (Figure 15A and B). Interestingly, 2.1% reactivation (Figure 15C) was observed after inducing differentiation to terminal blood lineages. We traced down this 2.1% reactivation via staining for lineage specific markers of CD14 (monocytes specific marker) and CD15 (neutrophil specific markers).

Specifically, the viral reactivation is in the CD14+CD15- monocytes fraction (Figure 15C). As a gate control, we performed similar assay to iPS cells without HIV integration.

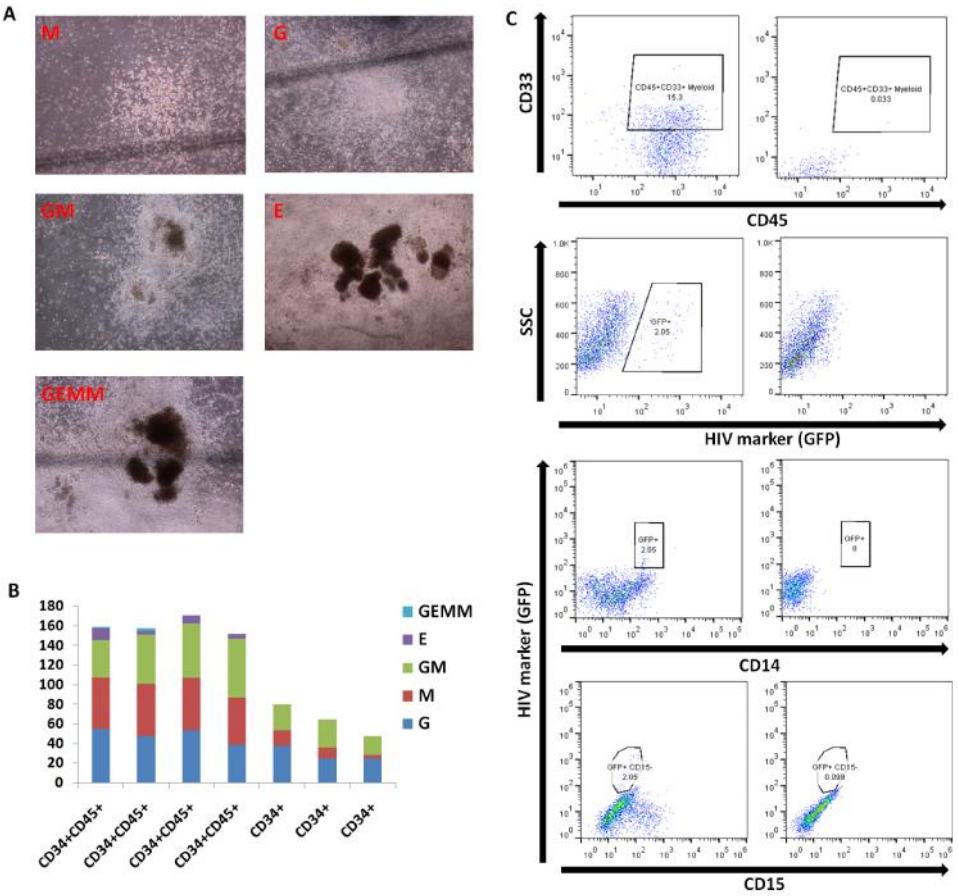

Figure 15. Induce differentiation of iPS colonies bearing HIV-Δenv latent proviral gene to blood lineage

(A). Multiple blood lineages after colony formation assay 14 days from the CD34+/CD45+ double positive cells generated by inducing differentiation of HIV-Δenv integrated iPS cells. G: Granulocytes. M: Macrophages. GM: Granulocytes and Macrophages mixed. E: Erythrocytes. GEMM: Granulocytes, macrophages and erythrocyte mix. (B). Colony counts and types observed after 14 days of CFU assay. Four independent CFU were performed to CD34+/CD45+ population while three independent CFU were performed to CD34+ only population. (C). FACS of multiple blood lineage markers and HIV-Δenv EGFP after collecting the cells from CFU assay. Left: HIV-Δenv integrated iPS induce differentiation with CD45 (myeloid lineage marker), CD33 (myeloid lineage marker), CD14 (monocytes markers), CD15 (neutrophil markers) stained. Right: control iPS line without HIV integration unstained.

There was evidence that CD34+ hematopoietic stem cells is able to silence the full-length HIV and potentially serves as a HIV reservoir (Carter CC, 2010). In contrast, since HIV-Δ*env*/*nef* is highly expressed in iPS colonies, we asked whether the latency phenotype of HIV in CD34+ population is also *nef* dependent. We induced differentiation of iPS

colonies with single insertion of HIV-Δ*env*/*nef* (iPS lateinfC4, Figure 9D) to CD34+ hematopoietic progenitor cells. We failed to obtain CD34+CD45+ population from this iPS line. However, we found that HIV-Δ*env*/*nef* expression has a 2.3-fold reduction upon inducing differentiation to CD34+ lineage (Figure 16), suggesting that CD34+ cells suppress HIV proviral expression partially via a Nef-independent mechanism.

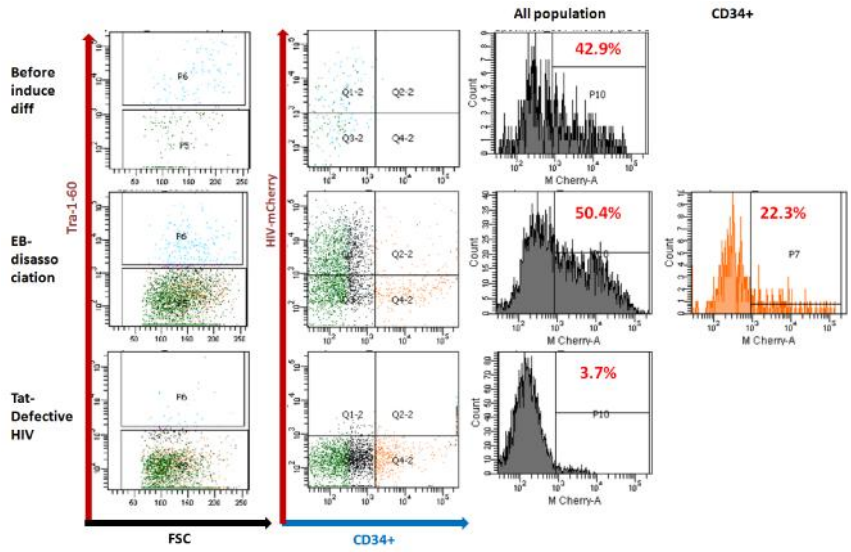

Figure 16. Inducing differentiation into CD34+ hematopoietic progenitor cells by EB differentiation. iPS colony with single integration of HIV-Δ*env*/*nef* (iPS late C3, Figure 9D) were differentiated into CD34+ cells. The iPS colony with a *tat*-defective HIV-Δ*env*/*nef* is used as a control (iPS C3).

Together, our result provides evidence Nef-induced HIV latency is more specific to human pluripotent stem cells, while inducing differentiation of iPS cells harboring a latent virus to both fibroblast and blood lineage can moderately reactive the proviral expression.

j. Identifying host factors that interact with Nef

As we observed a strong phenotype of Nef inducing viral latency in human pluripotent stem cells, to further identify the host factors involved in Nef's silencing of viral genes, we knockdown the host protein of N-myristoyltransferase 1 and 2 (NMT1 and NMT2)

that both are reported to interact with Nef to perform myristoylation(Seaton and Smith, 2008; Morgan et al., 2011; Jager et al., 2012). Other Nef partners, including ACOT8(Jager et al., 2012), PAK2(Rauch et al., 2008; Foster et al., 2011), Hck were also tested. After confirming successful knockdown of the host proteins in H9 cells, we co-infect the cells with HIV-Δ*env/nef* with pSIN-Nef or pSIN-EGFP (Figure 17) and checked the genomic copy number of HIV-Δ*env/nef* in both infections (Figure 18B).

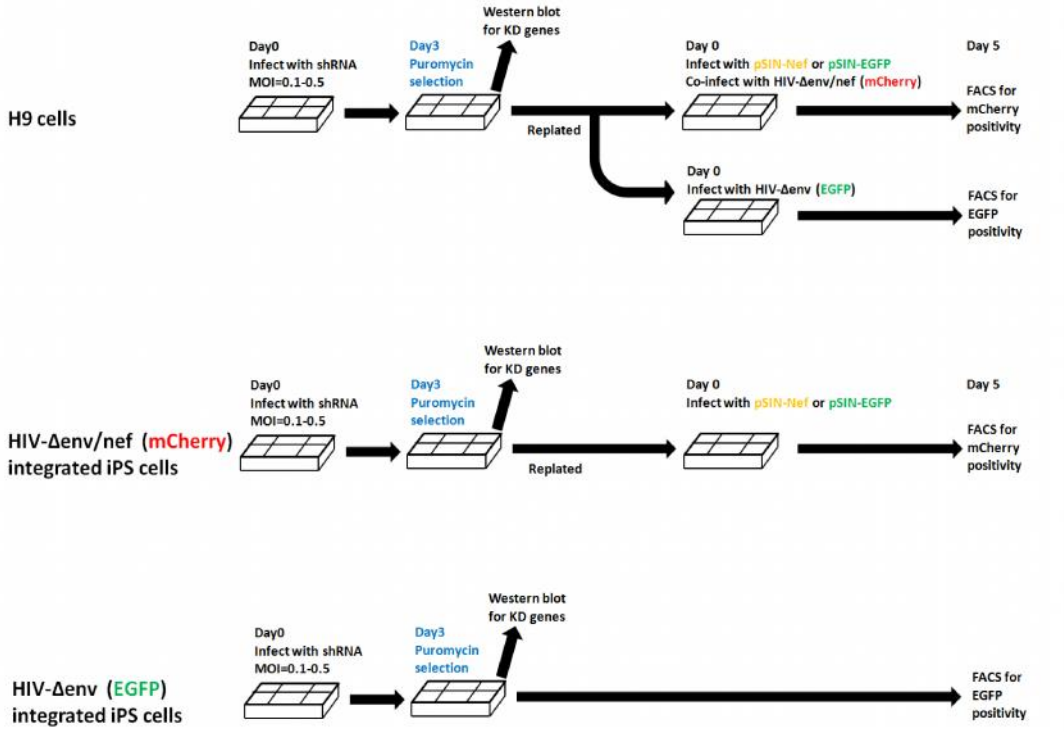

Figure 17. Strategic description of shRNA assay for Nef host partners.
(A). H9 cells were first infected with shRNAs encoded by pLKO lenti vectors (Sigma Mission). The cells were subjected to puromycin selection and western blots were performed to confirm the level of gene knockdown. The cells were replated for coinfection of HIV-Δ*env/nef* with pSIN-Nef or pSIN-EGFP, or infected with HIV-Δenv alone. Five days post infection, we surveyed the level of HIV expression and viral integration copy numbers. (B). The colonies with single integration of HIV-Δ*env/nef* were challenged with shRNAs to host Nef partners. After puromycin selection and western blots to confirm the knockdown level of host genes, we infected the cells with pSIN-Nef or pSIN-EGFP. Five days post infection, we surveyed the level of HIV expression and viral integration copy numbers. (C). The iPS colony bearing single integrated HIV-Δenv (Figure 9D) were challenged with shRNA targeting host Nef partners. After puromycin selection, the EGFP level for viral reactivation was measured by FACS.

We found that compared to random shRNA control, all the Nef partners reduced the effect of Nef in inducing viral silencing except PAK2 (Figure 18A). NMT1 and ACOT8,

which almost completely abolish Nef's silencing effect, were also reported to strongly interact with Nef in both 293T and Jurkat cells (Jager et al., 2012).

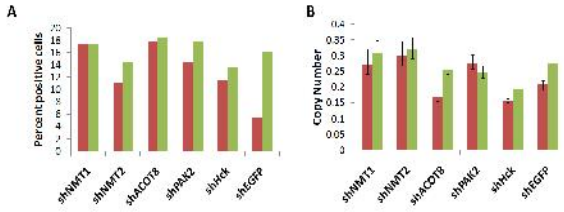

Figure 18. Knocking down Nef partners in pluripotent stem cells impairs Nef role of silencing HIV proviral genome in trans.
A. Percent of viral expressing cells measured by FACS in different shRNA groups after co-infection of HIV-Δ*env/nef* with pSIN-Nef (red) or pSIN-EGFP (green) B. The copy number measurements of HIV-Δ*env/nef* in pSIN-*nef* coinfection (red) or pSIN-EGFP coinfection (green).

II.5 Discussion

a. Model of HIV latency in human pluripotent stem cells

Previous studies reported the mechanism of mouse embryonic stem cells silencing MLV by forming histone modification machinery at the PBS region of MLV genome(Wolf D, 2007; Wolf D, 2008; Wolf D, 2009). However, in human pluripotent stem cells, after changing the PBS of MLV, we didn't observe reactivation of the MLV transcription. These results indicate that human pluripotent stem cells utilize a different system for retroviral silencing.

To further confirm the pluripotent state is needed to silence the virus and *nef* is responsible for inducing the latency, we performed reprogramming experiment and obtained single integrated HIV iPS colonies. The result is consistent with our previous observations. After supplementing Nef into the HIV- Δ*env/nef* infected human pluripotent stem cells, we observed silencing of HIV- Δ*env/nef* genome. These results suggest that the Nef is responsible for inducing the viral silencing.

To further analyze the mechanism of HIV latency in human-pluripotent stem cells, we performed ChIP-PCR assay and found that the HIV-Δenv genome is enriched for H3K9me3 marker while the HIV-Δ*env*/*nef* genome is enriched for H3K4me3. This result shows that the Nef induced HIV latency in human pluripotent stem cells is at epigenetic level.

We also performed differentiation induction assay towards blood or fibroblast lineages and showed limited reactivation of HIV- Δenv. However, when we infected the differentiated cells directly, the differentiated cells showed much less viral silencing compared to human pluripotent stem cells. These results suggest the model that pluripotency associated host factors are responsible for silencing the HIV but after it established long-term silencing in human pluripotent stem cells, the epigenetic modification of viral genome became irreversible so that the removal of the host pluripotent factors cannot reactivate the virus.

To try to identify the host factors responsible for viral latency, we knockdown five host genes that closely interact with Nef. The result showed that after knocking down these genes, the HIV silencing is compromised, indicating they play roles in inducing HIV latency in human pluripotent stem cells.

To summarize, our study suggested the model of HIV latency in human pluripotent stem cells shown below:

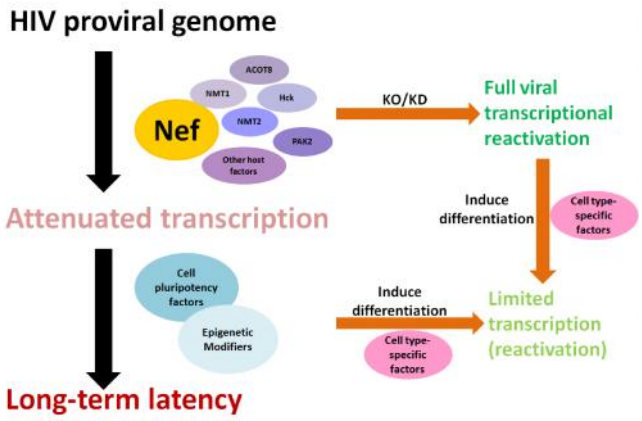

Figure 19. Suggested model of HIV latency in human pluripotent stem cells

b. Function of Nef protein

In our study, we found that HIV-1 Nef played important role in regulating HIV transcription in human pluripotent stem cells. This role of Nef has not been reported before.

Nef is a HIV-1 accessory protein that has 206 amino acids. In 1991 Kestler and colleagues reported that infecting rhesus monkeys with SIVmac239, a SIV virus strain that has a premature stop codon of Nef, results in attenuated viral replication *in vivo*(Kestler et al., 1991). It has also been reported that patients infected with HIV harboring defective Nef remained asymptomatic 14-18 years after HIV-1 infection without any HAART(Learmont et al., 1999). These *in vivo* observations indicated that *nef* played important role in viral replication and maintaining high viral load *in vivo*. However the molecular mechanism is poorly understood.

It is worth noticing that because *nef* has no catalytic activity, its pathogenetic function is mediated through cellular machinery or binding to HIV genome directly(Geyer et al., 2001; Malim and Emerman, 2008). In vitro experiments showed that Nef can bind to the LTR region of HIV genome and negatively regulates the transcription(Ahmad and Venkatesan, 1988). This observation was made in a cancer cell line SW480. People named this protein Nef due to the fact that it served as "Negative regulatory factor". Further study showed that *nef* can induced cell death in infected cells (Zauli G, 1999; Simmons et al., 2001) and induce apoptosis in bystander cells(Lenassi et al., 2010). However, people also observed that in T cells, HIV without Nef has reduced viral replication rate(Terwilliger E, 1986; Chowers MY, 1994; Spina CA, 1994; Pizzato M, 2008; Emert-Sedlak L, 2009; Ptak RG, 2010). Other study showed that Nef can prevent the cell death of infected cells by down-regulating cell surface MHCI, CD4, CD8, and CD28 and avoid visibility from immune cells (Leonard JA, 2011). These are conflicting observations, and our study showed that Nef reduced HIV transcription and induce cell death, which was consistent with the observation of Ahmad and Simmons. It is possible that Nef's function is largely dependent on cell type due to the fact that it has to collaborate with certain host factors. Those host factors' expression level and functions may differ in different cell types.

Nef function is largely dependent on myristoylation on the protein. N-myristoyltransferase-1 (NMT1) and N-myristoyltransferase-2 (NMT2) have been reported to directly myristoylate Nef protein though Nef was preferentially myristoylated

by NMT2 *in vitro*(Seaton and Smith, 2008; Jager et al., 2012). The myristoylation of Nef is required to activate cell signaling pathways including enhancement of p21-activated protein kinase (PAK2) autophosphorylation and interacting with Hck to Perturb the function of the Golgi complexes Hck(Hiyoshi et al., 2012). It is possible that in human pluripotent stem cells, players downstream of PAK2 and Hck interact with HIV transcriptional machinery to regulate HIV expression level, and Nef indirectly serve as transcriptional regulator by interacting with PAK2 and Hck.

c. HIV cytotoxicity in human pluripotent stem cells

We performed Nef supplementing experiments and observed suppression of HIV-Δ*env*/*nef* expression level in both co-infection and super infecting HIV-Δ*env*/*nef* iPS colonies (Figure 13), which is consistent with our previous observation of Nef's role in silencing the proviral genome in human pluripotent stem cells. However, we failed to observe the cytotoxicity effect after supplementing Nef (data not shown). We suspect the cytotoxicity that we observed in HIV-Δenv infection is likely not to be induced by Nef. Though previous arts reported Nef's role in inducing cell apoptosis via inducing the up-regulation of CD95/CD95L pathway (Zauli G, 1999), we suspect the in our experiment setting the cytotoxicity effect is not induced by Nef but the residual envelope expression (Westendorp MO, 1995).

Apoptosis, pyroptosis and necroptosis (non-apoptotic cell death pathway) can be induced by viral infection and serve as natural immunity to viral replication (Upton and Chan,

2014). Necroptosis can also be induced by NLRP1 and RIP3 in hematopoietic stem cells and embryonic stem cells (Kaiser et al., 2011; Masters et al., 2012). Pyroptosis is induced by NLRP3 and Caspase-1(Rathinam et al., 2012). Apoptosis is mediated by an extrinsic signal of FADD activating caspase-8(Kagi et al., 1994) or intrinsic signal of Apaf-1 activating caspase-9(Cain et al., 2000).

To further investigate the mechanism of HIV cytotoxicity in human pluripotent stem cells, we propose to use cell death inhibitors to further define how the cells died after HIV infection. Caspase-1 inhibitors (Millipore 400010) can block the pyroptosis(Thornberry et al., 1992). RIPK1 inhibitor (necrostatin-1, SIGMA N9037) and RIPK3 inhibitor are able to block necroptosis(Degterev et al., 2005). The Q-VD-OPh (SIGMA SML0063)(Caserta et al., 2003; Lavallard et al., 2009) and Z-VAD-FMK (SIGMA V116) (Hara et al., 1997) are broader caspase inhibitors thus can effectively block apoptosis and necrosis. We propose to screen the above chemicals and investigate which chemical is able to block the cell death upon HIV-Δenv infection in human pluripotent stem cells.

d. HIV long-term latency in human pluripotent stem cells

Though we identified multiple host proteins that can reduce the effect of Nef's role of silencing HIV provirus, we cannot reactivate the silenced HIV-Δenv in pluripotent stem cells via infecting shRNAs into iPS cells already bearing a HIV-Δenv latent gene. This result indicates that Nef functions at early stage after viral integration. The fact that infecting HIV-Δenv into H9 cells already selected for shRNAs also cannot reactivate the

HIV-Δenv led us to hypothesize that there are mechanism other than Nef that are responsive for the silencing of HIV-Δenv. We exclude the possibility of integrational sites selection bias by sequencing the viral integrational sites. We further hypothesize that cell death induced by HIV-Δenv probably exert additional effort on viral genome suppression through other intracellular pathways.

Another possibility is that HIV is silenced in human pluripotent stem cells via epigenetic modification such as CpG methylation. HIV long-term silencing via hyper-methylation on HIV proviral LTR has been reported to be associated with HIV long-term elite controller(Palacios et al., 2012). Other chromatin modification including histone deaceylation and histone methylation are both highly involved in the HIV latency (du Chéné I, 2007; Tyagi M, 2007; Pearson R, 2008; Archin NM, 2009; Friedman J, 2011). To further investigate the mechanism of HIV long-term silencing, we propose to conduct a bisulfite-sequencing to map the DNA methylation site on HIV genome. Furthermore, reactivation of the virus under extreme circumstances *in vivo* are important to restore viral replication restoring, and reactivating long-term reservoir by drugs can potentially lead to cure of AIDS. Here we propose to screen a library of DNA methylation inhibitors, HDAC inhibitors (such as TSA), 5-Azacytidine and Prostratin.

e. HIV latency in other cell lineages

We observed reactivation of HIV-Δenv in different cell lines after inducing differentiation. Since the expression level of *nef* partners in fibroblast and H9 cells are similar, according to public microarray data (GDS3892[ACCN]), the mechanism of viral

expression associated with cell pluripotency may not be regulated via expression control of Nef partners. We observed a bivalent histone status of HIV proviral genome in human pluripotent stem cells, which is enriched for both H3K9me3 and H3K4me3. Multiple genes in human stem cells maintain the bivalent histone modifications (Pan et al., 2007). We hypothesize that the difference of HIV expression in regards to cell pluripotency is regulated by global histone platform changes in cells when inducing differentiation. However, our observation shows that the reactivation effect is moderate by inducing differentiation, while directly infecting HIV-*nef* virus into cells with different pluripotency shows more pronounced differences in viral expression. This phenomenon indicates that after initial histone modification performed by players downstream *nef*, the virus genome is maintained in a bivalent status. This process is subjective to cell pluripotency. Further epigenetic modifiers are recruited afterwards that probably perform DNA modification to silence the virus in long-term, and the second modification process is not subject to cell pluripotency. Therefore, by inducing differentiation, the viruses cannot be reactivated due to a long-term silencing on HIV genome.

To better test whether the mechanism of HIV latency discovered using our model of human pluripotent stem cells can be transferable to human multipotent stem cells such as hematopoietic stem cells (HSC), the natural reservoir of HIV, more experiments such as infecting HSC with HIV-Δ*env*/*nef* and HIV-Δenv will be necessary.

Chapter III A Genome-wide association study reveals multiple new genes for host control of HIV-1 Subtype-C AIDS disease progression in an African population

III.1 Abstract

The Sub-Saharan African population infected with HIV-1C makes up the largest AIDS-patient population in the world and exhibits large heterogeneity in disease progression before initiating HAART. Previously, HLA-B*5701, other HLA alleles (Fellay et al., 2007; Colombo et al., 2008), Ring Finger Protein 39 (RFN39) and Zinc Ribbon Domain-Containing 1 (ZNRD1){Fellay, 2007 #1055} were identified as associated with VL set point or disease progression by genome wide association studies (GWAS) in HIV-subtype B infected European patients. However, little research has been done on control of HIV Subtype-C in the African population. Additionally, most previous studies to examine issues of progression in HIV/AIDS were based on single points in time for CD4 or viral load (VL) over time in treatment-naïve HIV infected patients. While searching for associations with trajectories of CD4 or VL are important, such indicators do not always follow linear trajectories. In our study, we performed an initial GWAS on 317 Tswana treatment-naive AIDS patients and replicated the result in a second independent cohort of 239. By applying a method of functional principal component analysis (fPCA) to characterize variability in CD4 and VL trajectories, we gained power with a relatively small data set. We identified 3 independent SNPs that are significantly associated with AIDS disease progression measured by combined CD4 and VL trajectories. The genes involved are HCG22, CCNG1 and ZBTB7C. The result is consistent after adjusting for sex and age. We also found several additional promising loci below significance level after using both fPCA and Linear mixed approximation for phenotypes: PRKAA2, ARHGEF12, TRIM26, ARL8B, AK095365, PCDH11X, and ZFYVE28.Our finding

raises the possibility that more SNPs linked to AIDS disease pathogenesis may yet be discovered if the approach used in this study is adopted and suggests that mechanistic studies of how CCNG1 and ZBTB7C interact with HIV-1 subtype C may uncover previously unknown pathways of HIV-host interaction. Collectively, though further replication studies are required with larger sample size to validate the finding, the genes we identified suggest new possibilities for examining biological mechanisms in HIV1-Subtype-C host control *in vivo*.

III.2 Introduction

Currently HIV-1C accounts for more than half of all global infections and several-fold more than any other subtype, including HIV-1B which predominates in Caucasian populations (Geretti, 2006). In 2012, of the 34 million people living with HIV globally, 23.5 million resided in Sub-Saharan Africa , where HIV-1 Subtype-C dominates (2012). In Botswana, the 2012 prevalence of HIV in adults (15-49 years) is estimated at 23.4% with 68.7% women (UNAIDS, 2013a). Given the diversity of the southern Africa population (Schuster et al., 2010), the fact that treatment-naive patients display significant heterogeneity in natural disease progression in our HIV-infected Botswana cohorts suggests that the host genetic makeup of individuals may contribute to their varying ability to maintain CD4 counts and viral load.

Studying the disease progression of AIDS is important for estimating optimal time to initiate Highly Active Antiretroviral Therapy (HAART) and understanding host restriction of viral replication, which may suggest avenues for development of drugs and vaccines. Despite the worldwide predominance of HIV-1 Subtype-C (HIV-1C), most

genome wide association studies (GWAS) have been conducted primarily on males of European ancestry infected by HIV-1 subtype B (Fellay et al., 2007; Fellay et al., 2009; van Manen et al., 2009; Troyer et al., 2011; van Manen et al., 2011; Bartha et al., 2013; McLaren et al., 2013). Previous GWAS for disease progression rate on HIV-subtype B infected European patients showed that most hits associated with viral load set points are on HLA-B*5701 and other HLA alleles (Migueles et al., 2000; Fellay et al., 2007; Dalmasso et al., 2008; van Manen et al., 2009; An and Winkler, 2010; O'Brien and Hendrickson, 2013). Other hits associated with disease progression measured by CD4 declining slope include ring finger protein 39 (RFN39) and zinc ribbon domain-containing 1 (ZNRD1) which encodes an RNA PolII subunit (Fellay et al., 2007). Notch4, C6orf12 and Trim10 were identified using the permutation procedure (Fellay et al., 2009).

However all these findings can only explain less than 10% of the differences in disease progression (Fellay et al., 2007), which calls for both a more accurate method for defining disease progression and a model combining the viral load with the CD4 trajectory. Using a single value of viral load set point or early CD4 counts cannot fully describe the course of measurements' change over time. Since we do not know in what manner SNPs might affect viral load or CD4 count, we employed functional principal component analysis (fPCA) (Gareth M. James, 2000; Yao F, 2005; Yao, 2007; Wu S, 2011) to characterize the variability in the trajectories of CD4 and VL. This fPCA method allows us to identify major modes of variation in the trajectories without any assumptions on the shape of the trajectories. It furthermore easily identifies typical modes of variation like mean and slope while also accommodating and identifying nonlinearity

in the trajectories. Compared to linear mixed simulation, our result showed that by adopting fPCA we can increase the study power and reveal additional SNPs that are associated with characteristics of CD4-VL combined trajectory other than slope.

In this study, we identified three significant host genetic restriction factors for viral replication using a novel definition of disease progression by fPCA. Our study is the first GWAS limited to HIV-1C and the southern African population. This is also the first study modeling AIDS disease progression trajectory using (fPCA) based on longitudinal observations of CD4 counts and viral load, and the first GWAS on HIV using Illumina Omni2.5 chips.

III.3 Material and Methods

a. Cohorts and patients eligibility

The BHP010 (Botsogo and Botsogo Extension Study) was a natural history observational prospective cohort study recruited from clinics in Gaborone, which was composed of HIV-1C infected individuals with starting CD4 cell counts above 400 cells per μl and not yet qualified for the Botswana HAART treatment program. Any one below 18, those with active AIDS defining illness requiring the initiation of HAART, the presence of an AIDS-related malignancy, or pregnant or breast feeding women who have been previously exposed to HAART were excluded. Participants returned to the clinic at 3-month intervals with an additional visit 1 month after enrollment and were followed up for up to 255 weeks (Figure 20A). The mean sampling time for CD4 was 91 weeks (Figure 20A). The end state of the study was censored (216/456), or enrollment on HAART treatment (120/456). Participants who could not be contacted after they have

missed three consecutive visits were considered “lost to follow-up” (90/456). Of the 456 patients enrolled, 370 were women and age distribution of BHP010 is shown in Figure 20B. DNA samples were collected at around three years post enrollment and 294 samples were available (Table 1). In this subset, the age and gender distribution are similar to the whole cohort, indicating a non-biased sampling (Table 2).

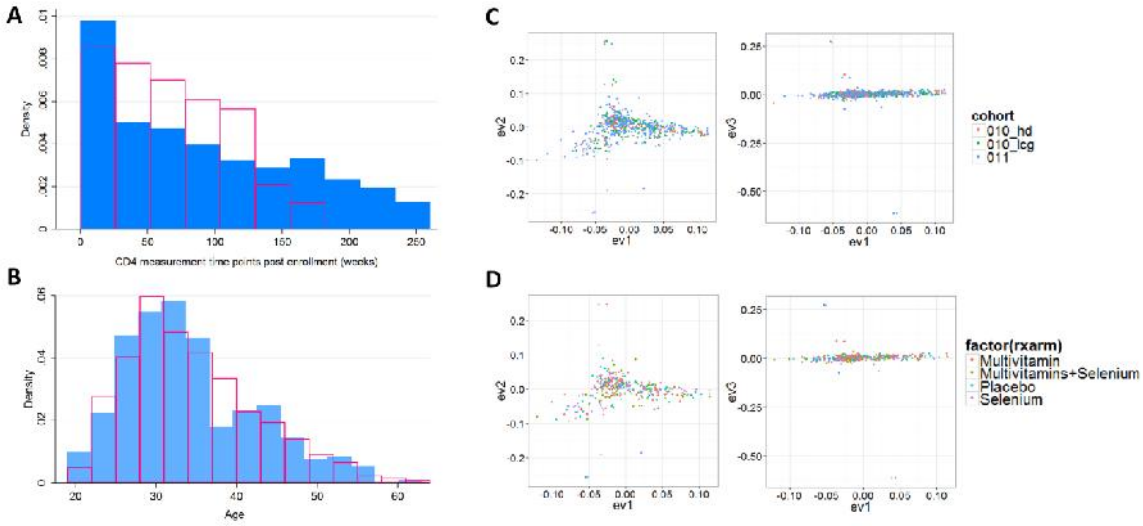

Figure 20. Population and sampling comparison between BHP010 and BHP011.
A. The histogram showing the distribution of CD4 sampling time point (weeks after enrollment) of BHP010 (blue) and BHP011(pink). BHP010 has a longer follow-up time with the longest of 255 weeks post enrollment and mean of 91 weeks. BHP011 has a shorter follow-up time with the longest of 169 weeks and mean of 56 weeks. B. Age distribution of BHP010 (blue) and BHP011 (pink). This figure involves all the patients in BHP010 and BHP011. C. EIGENSTRATE using genotypes of BHP010 and BHP011 after standard QC(Table 5). D. EIGENSTRATE using genotypes of BHP011, showing homogenous distribution of different treatment arms.

The BHP011 (Dikotlana Study) was a randomized, multifactorial, double-blind placebo-controlled trial to determine the efficacy of micronutrient supplementation (supplementation of multivitamin, selenium or both) in improving immune function and preventing early mortality of HIV-1C infection (Baum et al., 2013). It was composed of 875 treatment naive patients with CD4 higher than 350/μl. 638 out of 875 patients were women and age distribution of BHP011 is shown in Figure 20B. DNA samples were

collected at around two years post enrollment and 364 samples were available (Table 1). In this subset, the age and gender distribution were similar to the whole cohort, indicating a non-biased sampling (Table 3). Patients were followed up for maximum 169 weeks post enrollment (Figure 20A). They returned to clinics every 6 months to get CD4 samples. The mean CD4 sampling time was 56 weeks post enrollment (Figure 20A).

Table 1 Information of cohorts

	BHP010, natural history observational cohort	**BHP011, randomized multivitamin treatment cohort**	**Combined (number with DNA)**
participants	456 (294)	876 (364)	1332 (658)
Intervention	No	Four arms: Multivitamines alone: 219 (93) Selenium alone: 220 (88) Multivitamines+Selenium: 217 (94) Placebo: 219 (88)	
Women percentage	81.0% (84.7%)	72.9% (71.8%)	75.7% (77.6%)
Median baseline CD4 (cells/mm3)	468.5 (484.5)	422.2 (424.8)	439.1 (456.0)
Median baseline VL (Copies/ml)	12.9K (10.1K)	14.0K (14.5K)	13.6K (12.1K)

* Numbers in parentheses indicate result of patients with DNA available

Table 2 Average baseline and median CD4 and VL (BHP010)

BHP010	**whole cohort**	**Has DNA**	**qualified for GWAS**	**call rate pass through**	**Involved in GWAS analysis**
participants	456	294	258	250	240
average baseline CD4	513.6	520.9	532.1	534.5	533.1
Median baseline CD4	468.5	484.5	497	497.5	497.5
average baseline VL	56577	38743	37258	36952	35787
Median baseline VL	12900	10090	9565	8815	8255
average age	33.9	34.2	34.2	34.2	34.3
Women percentage	81%	85%	84%	83%	83%

Table 3 Average baseline and median CD4 and VL (BHP011)

BHP011	**whole cohort**	**Has DNA**	**qualified for GWAS**	**call rate pass through**	**Involved in GWAS analysis**

Table 3 Average baseline and median CD4 and VL (BHP011) (Continued)

participants	875	364	334	326	321
average baseline CD4	462.3	477.0	475.3	476.6	475.6
Median baseline CD4	422.2	424.8	427.8	432.4	429.7
average baseline VL	67600	72269	64972	66380	66189
Median baseline VL	14000	14500	14600	14900	15200
average age	34.0	35.0	35.3	35.2	35.5
Women percentage	73%	72%	71%	72%	71%

b. Genotyping process and quality control

The genotyping was performed using the protocol of Illumina® Infinium® Omni2.5 Multi-Use Assay. Due to discontinuation of manufacturer's product, two kinds of chips were used: Illumina HD BeadsChip and Illumina LCG BeadsChip. The numbers of patients genotyped by each kind of Illumina BeadsChip are shown in Table 4.

Table 4 Genotyping platforms used for BHP010 and BHP011

	Illumina HD	**Illumina LCG**
BHP010 (Botsogo)	112	142
BHP011 (Dikotlana)	0	335

Note: One patient from BHP010 was genotyped twice: one by HD Chip and the other by LCG Chip.

Samples with DNA concentration higher than 38ng/μl were qualified for genotyping. As shown in Table 5, of the qualified 592 samples, 4 samples didn't get genotyped due to discontinuation of manufacturer's product, so we genotyped 588 patients. Among them, 7 patients' call rate didn't pass 0.99. We also checked the specification of the gender to verify that patients with gender recorded in the clinical records as male should have heterozygous X genotypes (>=1%) or who if recorded as females should have

homozygous X genotypes. Six patients who didn't match the criteria were excluded. We also only include patients that have at least two clinical laboratory observations for both CD4 and VL; 19 patients were excluded by this criteria. After the initial QC, 556 patients were selected for further analysis, 239 of them are from BHP010 and 317 are from BHP011. Detailed characteristics of the patients before QC and after each step of initial QC are shown in Table 5.The gender ratio, median baseline VL, median baseline CD4, average baseline VL, average baseline CD4 and average age after each step of QC were all similar to the initial cohort for both BHP010 and BHP011, indicating unbiased sampling (Table 2, Table 3).

To avoid artifacts associated with either chemistry or clustering errors, we performed the quality control to the SNPs included in the association study as follows:

(1). Infinium BeadStudio Raw Data clustering:

Due to the incompatibility of LCG and HD chip data processing through Illumina Genome Studio, BHP010 on HD, BHP010 on LCG and BHP011 on LCG were processed separately. We used the reclustering file based on the previous unpublished Malawi genotyping data we generated. We applied the Quality control consistent with previous report (Fellay et al., 2007) . After the initial reclustering, any sample that had very low intensity or a call rate lower than 95% is deleted. All SNPs on Y chromosomes were individually evaluated to make sure they are only called in male samples. All SNPs on X chromosomes were also individually evaluated to make sure no male was heterozygous. The autosomal SNPs fulfilling any one of the five criteria were screened within BeadStudio and reclustered manually: call frequencies of less than 0.62; the mean of the normalized r-values for the AB genotypes of less than 0.28; the mean of the normalized

theta angles for the AB genotype is smaller than 0.2 or larger than 0.8; the Het Excess value is between -1.0 and -0.3 or between 0.2 and 1.0; and a Minor Allele Frequency less than 10%. If a SNP fails the reclustering, the SNP was zeroed manually. This process led to deletion of 3739, 6727, 6263 SNPs in BHP010 genotyped by Illumina HD BeadsChip, BHP010 genotyped by Illumina LCG BeadsChip and BHP011 genotyped by Illumina LCG BeadsChip, respectively.

(2). Minor Allele Frequency (MAF) check, genotyping missingness and Hardy-Weinberg Equilibrium (HWE):

After the reclustering, we further excluded SNPs that had Minor Allele Frequency lower than 5% in both BHP010 and BHP011. The rationale behind this quality control is that at least 28 people should bear the SNP in the cohort. We excluded SNPs with call rates smaller than 0.95 after reclustering. We further tested the Hardy-Weinberg Equilibrium with α=0.0001 and only included the SNPs that passed the test. The detailed result of the QC is shown in Table 5.

Table 5 Sample and SNP filtering for this study

A) Sample filtering		
Filtering	**Dropped***	**Included***
Total subjects		1332 (456, 876)
Blood sample collected	580 (162, 418)	752 (294, 458)
DNA sample available	94 (0, 94)	658 (294, 364)
Qualify for Genotyping	66 (36, 30)	592 (258, 334)
Genotyped	4 (3,1)	588 (255, 333)
Call rate pass 0.99	7 (5,2)	581 (250, 331)
Gender matched	6 (1,5)	575 (249, 326)
More than two clinical observations	19 (10,9)	556 (239, 317)
Total used		**556 (239, 317)**

* Numbers in parentheses indicate result of BHP010 (left) and BHP011 (right)

Table 5 Sample and SNP filtering for this study (Continued)

B) SNP filtering		
Filtering	**Dropped**	**Included**
Total number of SNPs on Illumina Omni2.5 BeadsChip HD and LCG		2454132
Consistent phenotype direction between primary and secondary cohorts for both CD4 and VL	680905	1773227
Call rate pass 0.99 in either cohorts	17799	1755428
MAF above cut off in either cohort	859517	895911
Pass HWE α=0.05	43022	852889
Autosomal	33857	819032
Total used (if including X and Y chromosomes)		852889

c. EIGENSTRAT method to perform population stratification

We applied the EIGENSTRAT as described before(Fellay et al., 2007). Basically we selected EIGENSTRAT axes to analyze the population ancestry. We selected SNPs based on the QC described above.

d. Functional principal component analysis steps

We performed the fPCA methods as described(Gareth M. James, 2000; James, 2002; Yao F, 2005) for log CD4 and log viral load in BHP010 and BHP011 separately. For each cohort and measure, we decomposed each person's trajectory into a mean function (shared among all patients) and a linear combination of Eigenfunctions, $\phi_k(t), k = 1, 2, 3$. Because the Eigenfunctions are constructed such that the first Eigenfunction explains the most variability in the data, the second Eigenfunction explains the second

most variability, and so on, only 3 Eigenfunctions are necessary to explain more than 99% of the variation in the data. The coefficients, or scores, ξ_{ik}, associated with the Eigenfunctions succinctly summarize the patient-specific trajectories.

$$\log \text{CD4}_{ij} = \mu^{(\text{CD4})}(t_{ij}) + \sum_{k=1}^{3} \xi_{ik}^{(\text{CD4})} \phi_k^{(\text{CD4})}(t_{ij}) + \epsilon_{ij}^{(\text{CD4})}$$

$$\log \text{VL}_{ij} = \mu^{(\text{VL})}(t_{ij}) + \sum_{k=1}^{3} \xi_{ik}^{(\text{VL})} \phi_k^{(\text{VL})}(t_{ij}) + \epsilon_{ij}^{(\text{VL})}$$

where for the ith patient at time t_{ij}, CD4_{ij} represents their observed CD4 count; $\mu^{(\text{CD4})}(t_{ij})$ represents the population mean function for log-CD4 evaluated at t_{ij}; $\phi_k^{(\text{CD4})}(t_{ij})$ represents the kth Eigenfunction evaluated at t_{ij}; $\xi_{ik}^{(\text{CD4})}$ represents the random score for the kth Eigenfunction, which can also be seen as a random effect; and $\epsilon_{ij}^{(\text{CD4})}$ represents random measurement error. Quantities are defined analogously for viral load.

The population means $\mu^{(\text{CD4})}$ and $\mu^{(\text{VL})}$ are estimated via nonparametric smoothing, and the Eigenfunctions are obtained via an eigendecomposition of the smoothed covariance matrices of log CD4 and log VL (Rice, 1991; Hall, 2006). To estimate the scores, we used the best linear unbiased predictor(Robinson, 1991) because the data on CD4 and viral load were not densely observed. Thus, to estimate the scores for person i, we borrow information from all other subjects. Only three Eigenfunctions were needed to explain more than 99% of the variability in the trajectories.

e. Linear regression model to characterize CD4 and VL trajectories

We also fit a more familiar linear mixed effects model (Laird, 1982; Smith et al., 2010; Zhang et al., 2010) to bolster the evidence from the FPCA model

$$\log \mathrm{CD4}_{ij} = \beta_0^{(\mathrm{CD4})} + \beta_1^{(\mathrm{CD4})} t_{ij} + \xi_{i0}^{(\mathrm{CD4})} + \xi_{i1}^{(\mathrm{CD4})} t_{ij} + \epsilon_{ij}^{(\mathrm{CD4})}$$

$$\log \mathrm{VL}_{ij} = \beta_0^{(\mathrm{VL})} + \beta_1^{(\mathrm{VL})} t_{ij} + \xi_{i0}^{(\mathrm{VL})} + \xi_{i1}^{(\mathrm{VL})} t_{ij} + \epsilon_{ij}^{(\mathrm{VL})}$$

This model corresponds to pre-specifying the model above to have $\mu^{(\mathrm{CD4})}(t_{ij}) = \beta_0^{(\mathrm{CD4})} + \beta_1^{(\mathrm{CD4})} t_{ij}$, $\phi_1^{(\mathrm{CD4})}(t_{ij}) = 1, \phi_2^{(\mathrm{CD4})}(t_{ij}) = t_{ij}$, and $\phi_3^{(\mathrm{CD4})}(t_{ij}) = 0$ (and likewise for viral load).

f. Testing

To characterize the strength of association between a SNP, G_i, and the trajectories of log CD4 and log viral load, we obtained variance component score statistics, similar to those used previously(Lin, 1997; Wu et al., 2011), that correspond to the working models

$$\xi_{ik}^{(\mathrm{CD4})} \sim N(\gamma_k^{(\mathrm{CD4})} G_i, \tau^2), \xi_{ik}^{(\mathrm{VL})} \sim N(\gamma_k^{(\mathrm{VL})} G_i, \tau^2)$$

where $\gamma_k^{(\mathrm{CD4})}$ represents the effect of the SNP on the trajectory of CD4. For the null hypothesis of no association between the SNP and the longitudinal trajectories of CD4 and viral load, H0: $\gamma_k^{(\mathrm{CD4})} = \gamma_k^{(\mathrm{VL})} = 0$, the form of the test statistic is given by

$$\sum_{k=1}^{3} \sum_{l=\mathrm{CD4,VL}} (n^{-1/2} \sum_{i=1}^{n} \hat{\xi}_{ik}^{(l)} G_i)^2$$

where $\hat{\xi}_{ik}^{(l)}$ is the best linear unbiased predictor of $\xi_{ik}^{(\mathrm{l})}$. To adjust for age and sex, the test statistic becomes

$$\sum_{k=1}^{3} \sum_{l=\mathrm{CD4, VL}} (n^{-1/2} \sum_{i=1}^{n} (\hat{\xi}_{ik}^{(l)} - \hat{\alpha}_{1k}^{(l)} \mathrm{age}_i + \hat{\alpha}_{2k}^{(l)} \mathrm{sex}_i) G_i)^2$$

where $\hat{\alpha}_{1k}^{(l)}$ and $\hat{\alpha}_{2k}^{(l)}$ are obtained under the null hypothesis.

Under the null hypothesis, the test statistic is distributed asymptomatically as a mixture of χ^2s. We obtain p-values by estimating its null distribution and comparing it to the observed test statistic. The overall error rate was controlled using the Benjamini-Hochberg false discovery rate (FDR) controlling procedure(Benjamini, 1995). The FDR was controlled at the level 0.10.

III.4 Results

a. Population characteristics and genetic stratification

Two independently recruited Tswana cohorts were involved in our study: a natural history observational prospective cohort (BHP010) observed prior to drug intervention, and a randomized multivitamin treatment cohort (BHP011) with four treatment arms (Table 1, also see Methods). Both cohorts consist of AIDS patients from Botswana that are infected with subtype-C and have not been exposed to HAART. Age distribution was similar for the two cohorts (Figure 20), despite BHP010 having a slightly higher proportion of women (Table 1). Based on EIGENSTRAT population stratification analysis(Price et al., 2006), the genetic makeup of BHP010 and BHP011 is not separated and the two genotyping platforms showed similar results (Figure 20C). The four treatment arms of BHP011 are also genetically similar (Figure 20D). After going through a stringent exclusion criteria (See methods, Table 5), 556 patients were finally involved in our association analysis.

b. Phenotype definition using Functional Principal Component Analysis (fPCA) to reconstitute of CD4 and VL trajectory

The natural history of treatment naive patients from BHP010 and BHP011 cohorts shows various disease progression patterns (Figure 21). We performed fPCA {Holte, 2012 #3311} separately on BHP010 and BHP011 due to the fact that the length of follow-up differed between the two cohorts. The most important modes of variation in the trajectories were selected such that 99% of the observed variability in the data was explained. This is analogous to regular principal component analysis where the most important directions of variation of a matrix are obtained via Eigen decomposition. These three modes highlight three typical trajectories of maintaining, declining and varying CD4/VL counts (Figure 21A). Extremity of the first function scores is able to separate patients maintaining high or low CD4/VL, while extremity of second function scores can separate flat versus declining CD4/VL scores (Figure 21B).The third Eigenfunction captures the variation of longitudinal trajectories. We then used the score of the first Eigenfunction to approximate the level of baseline CD4/ VL, and the score of the second Eigenfunction to approximate the progressiveness of their diseases.

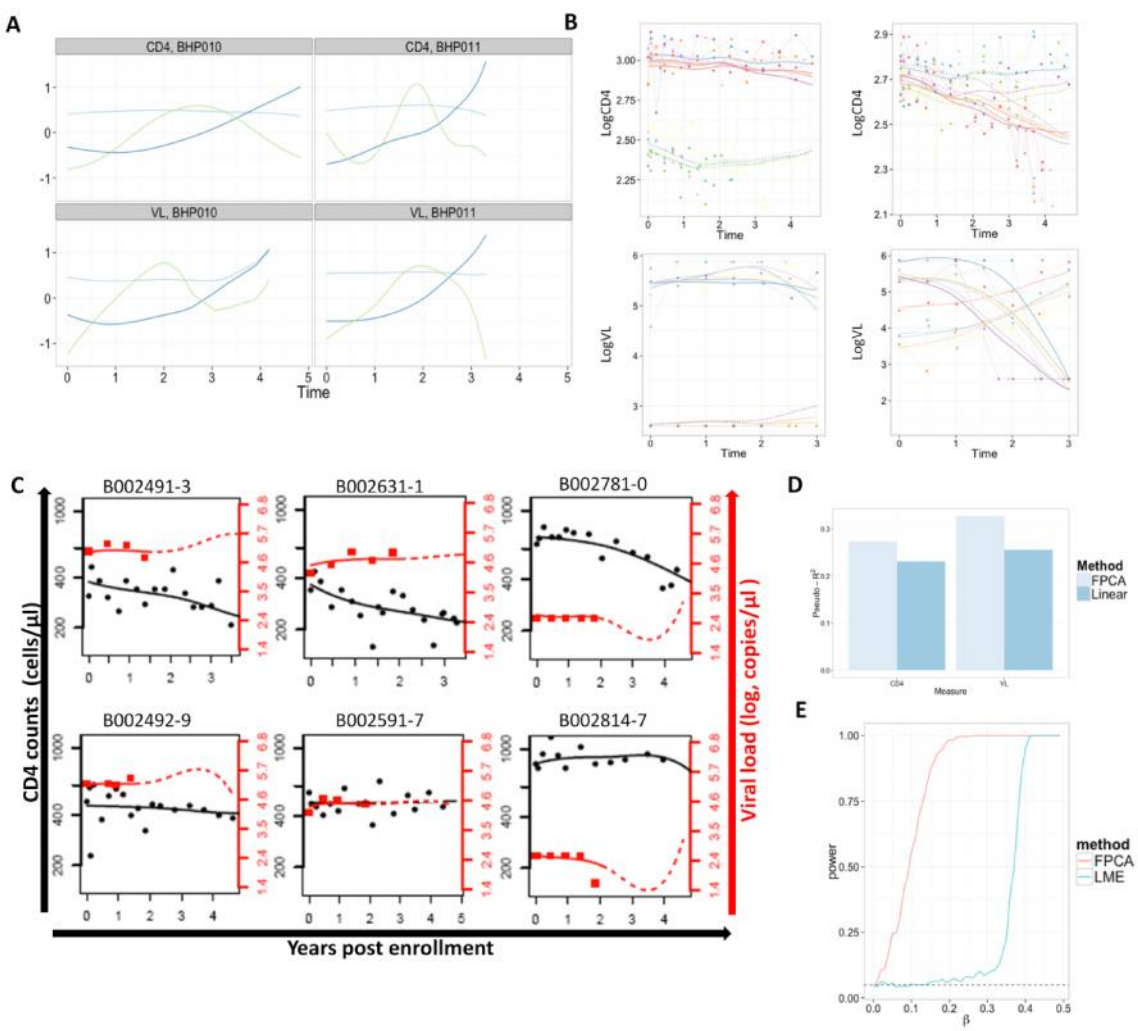

Figure 21. Functional principle component analysis to estimate Eigenfunctions of CD4 and VL trajectories.

A. Estimated Eigenfunctions used to fit CD4 and viral load longitudinal data for BHP010 and BHP011 separately. Three functions were used for viral load and CD4 separately in each cohort. B. Extremity of the first Eigenfunction coefficients (Left) separate patients with high or low CD4 counts (Upper) or Viral load (lower). Extremity of the second Eigenfunction coefficients (Right) separate patients with fast declining CD4 or VL versus flat or increased CD4 or VL. C. Examples of Eigenfunction fit for six patients. With similar viral load counts (red), patients showed different patterns of CD4 trajectory (black) over time. Solid lines represent approximation of CD4 or VL trajectories by fPCA. Dash lines represent predicted trajectories of VL by fPCA. D. Proportion of variation explained by FPCA (light blue) and linear model (dark blue) as measured by pseudo-R^2. Larger values indicate better model fit to the data. E. Power curve of association study using FPCA (red) or linear regression (blue).

We found that by fitting curves via fPCA Eigenfunctions, we can largely reconstitute the CD4 and viral load trajectories for patients (Figure 21C) with a lower variation represented by higher pseudo-R^2 compared to linear regression (Figure 21D). FPCA provides relevant phenotypes by separating fast progressors from slow progressors

(Figure 21B, Figure 22) and it does not require the restrictive linearity assumption of linear mixed effects models, potentially gaining power to detect effects on nonlinearity in CD4 and VL trajectories. .Additionally, in order to account for the profound interaction between viral load and CD4, we combined VL and CD4 fPCA scores (coefficients) to define the phenotype of patients.

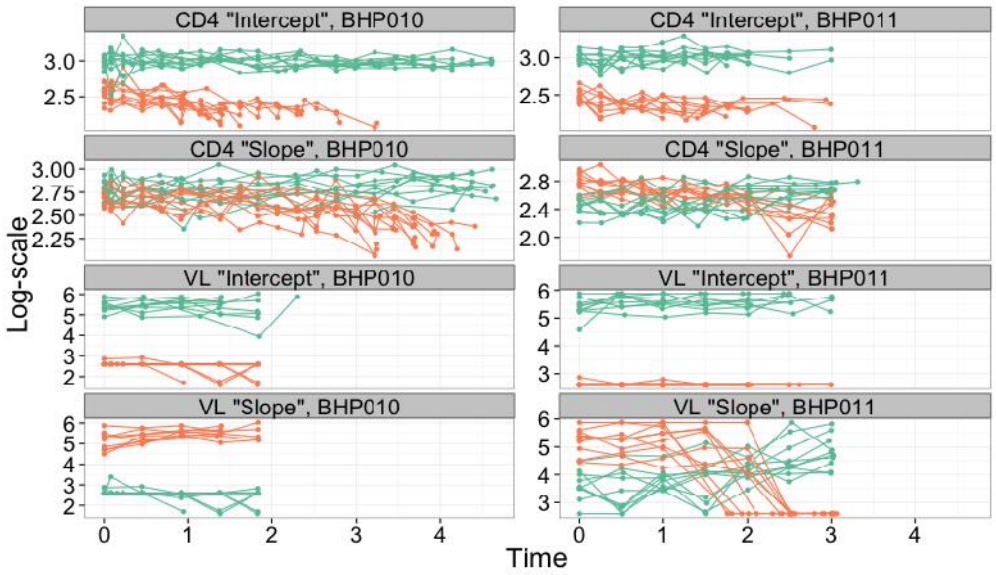

Figure 22. Extremity of intercept and slope values using linear model
Green: high intercept value or low absolute value of slope. Red: low intercept value or high absolute value of slope. The intercept extremities separate patients with high viral load or CD4 (green) versus patients with low viral load or CD4 (red). The slope extremities separate patients that are fast progressors (red) versus slow progressors (green).

c. GWAS using BHP011 as primary cohort and BHP010 as replication cohort

We conducted an initial GWAS on BHP011 sized 317 as primary cohort with 819032 SNPs included (Table 5) and selected 61 SNPs with P-value lower than 6×10^{-4} to replicate in the second cohort of BHP010 sized 239. After false discovery rate (FDR) adjustment with the significance level of 0.1 we identified 39 SNPs that stand

significantly associated with HIV disease progression with a consistent association direction in two cohorts. We tested the direction of the effect and eliminated SNPs that were significant after replication but have opposite direction of effect in the primary cohort and the secondary cohort (Figure 23, Table 6). After Linkage disequilibrium analysis, we found 8 SNPs lying in independent regions on the human genome (Table 7, Figure 24) that have same association direction in both cohorts. Among them, 3 reached genome-wide significance level with a P-value lower than 2.38×10^{-6} in a pooled analysis of the two cohorts (Table 8, Figure 24 A and B). Three genes are involved in these regions which include HCG22 (peak variant rs2535307, Pcombined=3.72×10^{-7}),CCNG1 (peak variant kgp22385164, Pcombined=1.88×10^{-6}) and ZBTB7C (peak variant kgp7815029, Pcombined=1.12×10^{-6}).We also found 5 more SNPs below genome-wide significant level in pooled analysis that passed the false discovery rate adjustment in replication. The genes involved are PRKAA2, ARHGEF12, AK095365, ZFYVE28 (Table 7). Additionally, we found 2 more SNPs (close to genes of IPO8 and CYP1B1-AS1) reached genome-wide significance level in combined analysis but they are only significant in one cohort and one SNP lying in the intron of DPYD that is significant in the primary cohort but didn't reach MAF 5% in the replication cohort (Figure 24B,Table 9).

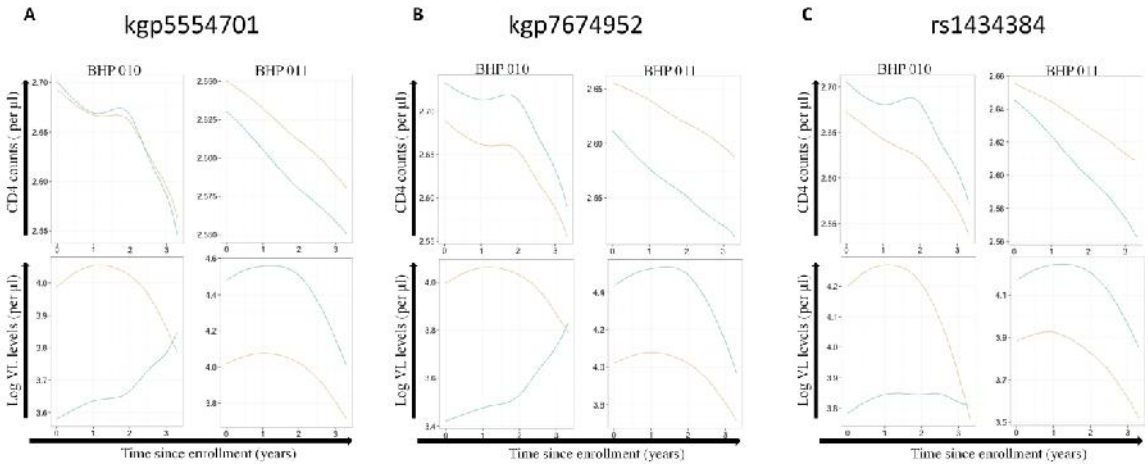

Figure 23. Average of trajectory of significant SNPs that are eliminated due to opposite direction in primary and secondary cohorts generated by fPCA model.
Yellow: The average trajectory of patients with 0 minor allele. Green: The average trajectory of patients with 1 minor allele

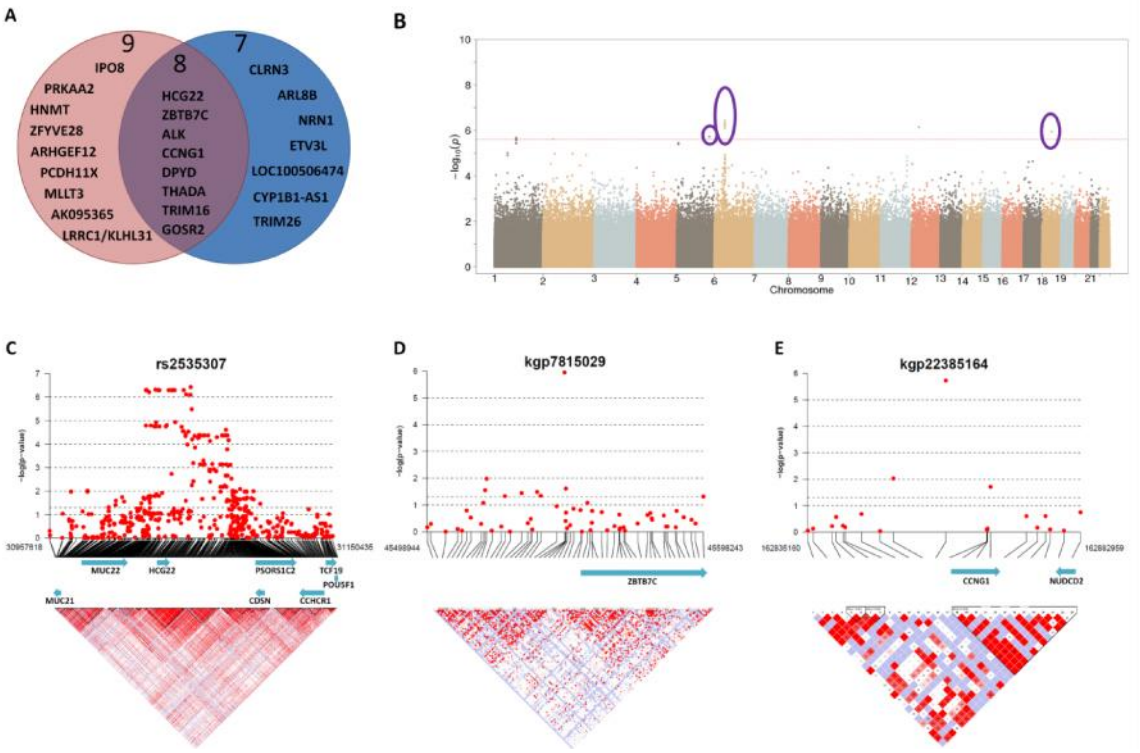

Figure 24. GWAS results at 852889 SNPs with using combined VL and CD4 fPCA scores as phenotype under a dominant model.
A. Genes involved in the top 100 SNPs (smallest genome-wide $P_{combined}$) identified using fPCA model (red) or linear mixed model (Blue). B. The Manhattan plot before adjusting for age and sex shown in combined p value of BHP010 and BHP011. Six SNPs in independent regions reached genome wide significance level. Three SNPs (in purple circle) are significant in primary cohort of BHP011 and remain significant after replication in BHP010. For the other three SNPs (above significance level but not circled), though their combined p-value reached significance level, they are significant in only one cohort but not the other. C-E. Linkage disequilibrium around significant SNP loci identified by fPCA model.

Table 6 Significant SNPs that have opposite effects on BHP010 and BHP011

SNP	locus	Annotation	Nearest Gene	Model	MAF (primary)	P-value (replicate)	P-value (primary)	P-value(combined)
rs1434384	4q32.1	14kb upstream	LOC340017	Linear/ fPCA	0.446875	0.00047	2.43E-05	2.21E-07
kgp5554701	3q27.3	intergenic		Linear/ fPCA	0.315625	0.008044	4.02E-06	5.90E-07
kgp7674952	4q22.1	intron	CCSER1	Linear/ fPCA	0.06875	0.002022	5.75E-05	1.97E-06
rs6956267	7p21.3	25kb downstream	ARL4A	fPCA	0.159375	0.012766	5.43E-05	1.05E-05
kgp5652371	4q21.23	intron	BC005018	fPCA	0.107813	0.003995	0.00029	1.70E-05
rs13165594	5q34	intergenic		fPCA	0.1	0.075816	1.81E-05	1.99E-05
kgp20189754	8q11.21	100kb downstream	UBE2V2	fPCA	0.075	0.039468	4.79E-05	2.68E-05
kgp1807804	3p25.2	intron	SYN2	fPCA	0.107813	0.006795	0.00029	2.79E-05
kgp22752575	16p13.3	intron	SRL	fPCA	0.21875	0.056438	3.55E-05	2.83E-05
kgp6123710	4p15.32	2.5kb upstream	C1QTNF7	fPCA	0.134375	0.022793	9.68E-05	3.10E-05
kgp9310091	1q43	intron	AKT3	fPCA	0.228125	0.077378	2.87E-05	3.12E-05
kgp13912306	17p12	intergenic		fPCA	0.0625	0.014882	0.000201	4.10E-05
kgp12811778	11p14.2	intron	BBOX1	fPCA	0.103125	0.015893	0.000198	4.30E-05
kgp11422540	17q24.2	intron	APOH	Linear	0.066532	0.047115	2.55E-05	1.77E-05
rs4803449	19q13.2	intron	AXL	Linear	0.454259	0.00296	0.000577	2.44E-05
kgp22814565	Xp21.3	10kb downstream	AK057304	fPCA/ Linear	0.06875	0.007055	3.52E-05	4.03E-06

Table 7 SNPs significant* after replication using fPCA model (excluding X and Y chromosomes)

SNP	locus	Annotation	Nearest Gene	Allele (ref/ alt)	MAF (primary)	P-value (replicate)	P-value (primary)	P-value(combined)
rs2535307	6p21.33	25kb downstream	HCG22	T/C	0.480456	2.39E-05	0.00083	3.72E-07
kgp7815029	18q21.1	20kb upstream	ZBTB7C	T/C	0.154688	0.064199	9.90E-07	1.12E-06
kgp22385164	5q34	5kb upstream	CCNG1	T/G	0.082813	0.006201	1.78E-05	1.88E-06
rs4912408	1p32.2	intron	PRKAA2	T/C	0.209375	0.00583	0.000109	9.73E-06
rs10892573	11q23.3	intron	ARHGEF12	T/C	0.141509	0.003182	0.000303	1.43E-05
kgp18828163	12p13.32	intron	AK095365	T/C	0.126563	0.037386	5.62E-05	2.96E-05
kgp11018583	2q31.1	intergenic		T/C	0.265625	0.005711	0.000445	3.53E-05
kgp3432145	4p16.3	exon	ZFYVE28	A/G	0.063091	0.006051	0.000525	4.34E-05

* SNP p-value lower than 6×10^{-4} inBHP011 (primary cohort) and passed FDR (α=0.1) in BHP010 (second independent cohort)

Table 8 Genome-wide significant loci that are significant in both cohorts by fPCA and Linear mixed phenotype simulation

SNP	locus	Annotation	Nearest Gene	Allele (ref/alt)	MAF (primary)	P-value (replicate)	P-value (primary)	P-value(combined)	Phenotype simulation model
rs2535307	6p21.33	25kb downstream	HCG22	T/C	0.480456	2.39E-05	0.00083	3.72E-07	fPCA
kgp7815029	18q21.1	5kb downstream	ZBTB7C	T/C	0.154688	0.064199	9.90E-07	1.13E-06	fPCA
kgp22385164	5q34	5kb upstream	CCNG1	T/G	0.082813	0.006201	1.78E-05	1.88E-06	fPCA
rs6921557	6p21.33	20kb upstream	HCG22	T/C	0.478056	1.26E-04	3.28E-06	5.37E-09	Linear
kgp7815029	18q21.1	5kb downstream	ZBTB7C	T/C	0.154688	0.065581	6.02E-06	6.22E-06	Linear

Table 9 SNPs significant in one cohort that fail to replicate as significant in second cohort

SNP	locus	Annotation	Nearest Gene	Model	MAF (BHP011)	Significant cohort	P-value in significant cohort	P-value in the other cohort	P-value(combined)
rs11050931	12p11.22	100kb downstream	IPO8	FPCA/Linear	0.384375	BHP010	7.13E-08	0.555184	7.15E-07
rs7606492	2p22.2	25kb downstream	CYP1B1-AS1	FPCA/Linear	0.459375	BHP010	2.90E-07	0.434883	2.13E-06
rs6703307	1p21.3	intron	DPYD	FPCA	0.064062	BHP011	2.09E-06	MAF<0.05	2.09E-06
kgp1867220	5p15.33	intergenic		Linear	0.315625	BHP010	3.62E-05	0.003608	2.20E-06

d. Adjusting by age and sex

We also adjusted the result by age and sex. All the SNPs discussed above still remain in the top 100 lowest by p-value after adjusting for sex and age (Figure 25). We excluded sex chromosome in all the analysis to avoid potential bias after adjusting for age and sex.

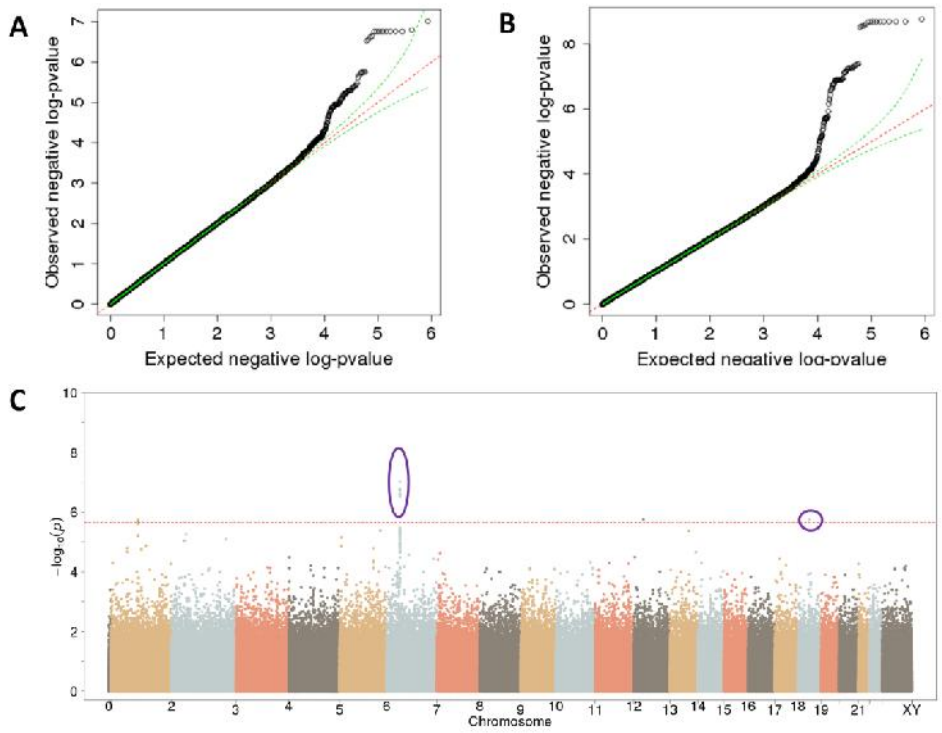

Figure 25: The QQ plot and Manhattan plot after adjustment by age and sex if including sexual chromosome in association analysis (dominant model) for fPCA.
A. QQ plot using fPCA model after adjusting for age and sex, including X and Y chromosomes. B. QQ plot using linear model after adjusting for age and sex, including X and Y chromosomes. C. Manhattan plot using fPCA model after adjusting for age and sex, including X and Y chromosomes. Top SNPs identified before stratification still remain top. Four independent SNPs passed significant threshold and HCG22 is the only one with consistent small p-value in both cohorts. All sex chromosome SNPs failed to reach the genome-wide significant level after gender and age adjustment.

e. Using the dominant model

Our results are based on the dominant model(Bush and Moore, 2012). We chose to use the dominant model because most autosomal genetic traits are inherited in a dominant pattern. However there are also autosomal traits inherited by non-dominant patterns. So we replicated our study using a non-dominant model to see whether we can gain power for some loci. We found that using the non-dominant model decreased power significantly, but SNPs with the lowest p-values remain the same as with the regular model (Figure 26).

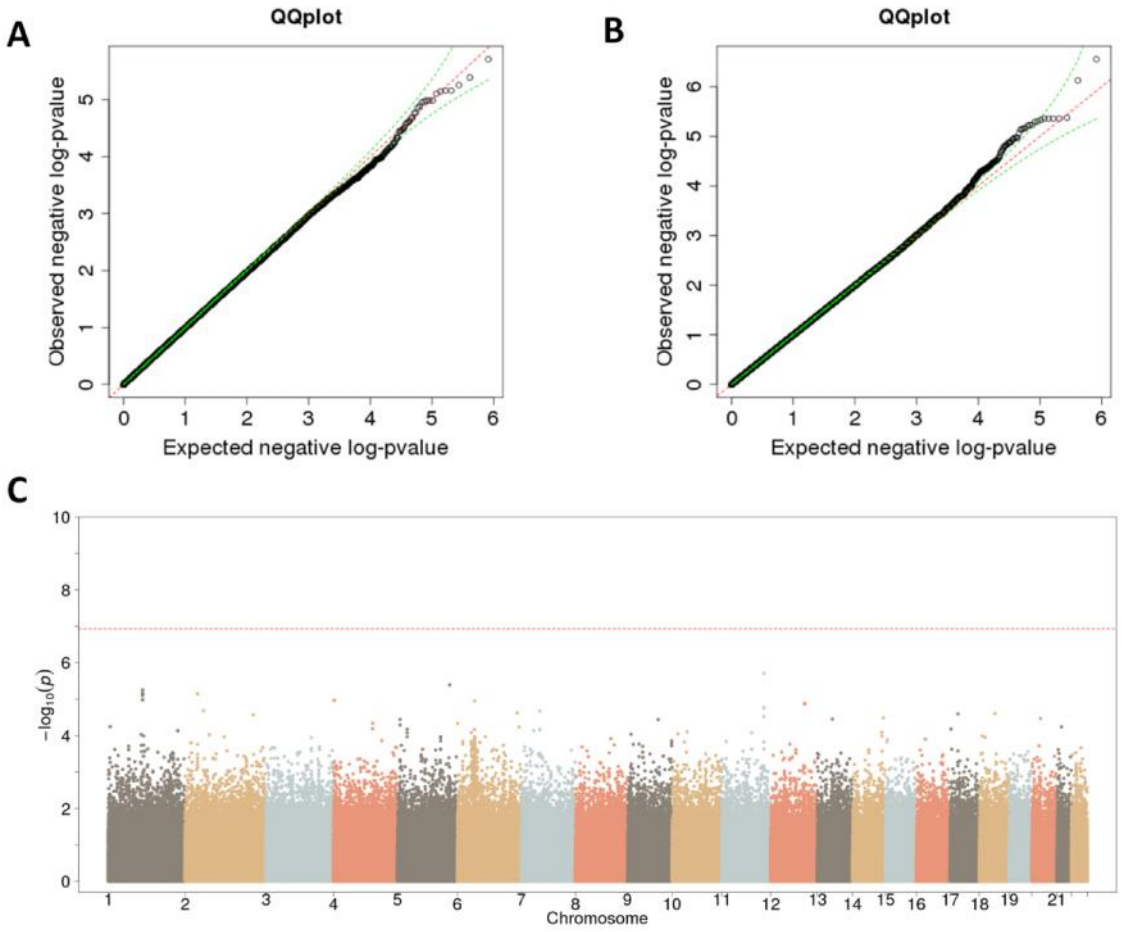

Figure 26. QQ plots and Manhattan plots using non-dominant model before stratifying by age and sex.
A. QQ plot using FPCA model before stratifying by age and sex. B. QQ plot using linear model before stratifying by age and sex. C. Manhattan plot using FPCA regression model before stratifying by age and sex. No SNP is significant in the non-dominant model.

f. Linear model for phenotype approximation

We also replicated the above association study using a traditional linear mixed model to estimate the CD4 and VL trajectories. Without adjusting for age and sex, we obtained 80 significant SNPs after FDR (a=0.1) adjustment lying in 5 independent regions of the human genome. These SNPs were in HCG22 (peak variant rs6921557, Pcombined=5.37×10^{-9}), ZBTB7C (peak variant kgp7815029, Pcombined=6.22×10^{-6}),

CCNG1(peak variant kgp22385164, Pcombined=9.14×10^{-6}) ,TRIM26 (peak variant kgp10188320, Pcombined=1.45×10^{-5}) and ARL8B (peak variant kgp8910014, Pcombined=1.99×10^{-5}) (Figure 27). Only HCG22 and ZBTB7C reached genome-wide significance level after FDR(a=0.1, p-value lower than 7.48×10^{-6}) (Table 8). It is worth noticing that though genes with the smallest p-values largely overlap in results obtained using the fPCA model and linear mixed model (Figure 24A), the linear mixed model has less power reflected by larger SNPs p-values for the same SNPs compared to the fPCA model, except for HCG22 (Table 7, Table 10). The result using the linear mixed model further confirms the association we observed using the fPCA model. Though the linear mixed model gives us a much lower p-value for HCG22, we generally gained power to detect other significant SNPs using the fPCA model.

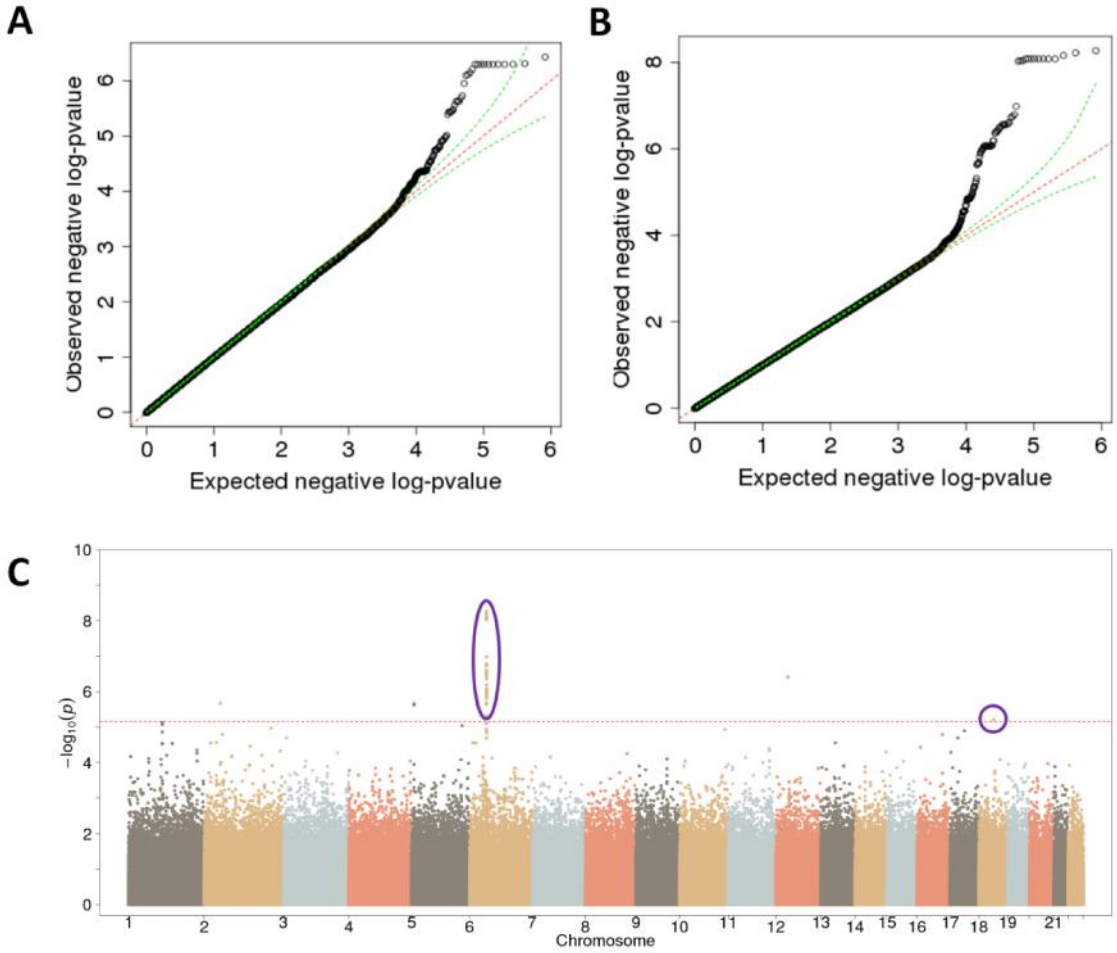

Figure 27. QQ plots and Manhattan plots using different phenotype simulation models before adjusting for age and sex (excluding sexual chromosome).
All of the plots used the dominant model. A. QQ plot using fPCA model before adjusting for age and sex. B. QQ plot using linear model before adjusting for age and sex. C. Manhattan plot using linear mixed model before adjusting for age and sex.

Table 10 SNPs significant* after replication using linear mixed model (excluding X and Y chromosomes)

SNP	locus	Annotation	Nearest Gene	Allele (ref/alt)	MAF (primary)	P-value (replicate)	P-value (primary)	P-value(combined)
rs6921557	6p21.33	20kb upstream	HCG22	T/C	0.480456	1.26E-04	3.28E-06	5.37E-09
kgp7815029	18q21.1	5kb downstream	ZBTB7C	T/C	0.154688	0.070688	6.03E-06	6.22E-06
kgp2238564	5q34	5kb upstream	CCNG1	T/G	0.082813	0.011275	9.25E-05	9.14E-06
kgp10188320	6p22.1	intron	TRIM26	T/C	0.440625	0.065684	1.74E-05	1.45E-05
kgp8910014	3p26.1	50kb upstream	ARL8B	A/G	0.1125	3.22E-03	0.000173	1.99E-05

* SNP p-value lower than 6×10-4 in BHP011 (primary cohort) and passed FDR (α=0.1) in BHP010 (second independent cohort)

g. Phenotype of patients with and without minor alleles

We further examined the CD4 and VL trajectories in patients with different variations of the significant SNPs. Of the three significant SNPs identified by the fPCA model, the trajectories of patients without minor alleles and with at least one minor allele display significant differences (Figure 28, Figure 29, Figure 30). This effect is reflected by the most significant difference in their first Eigenfunction score and moderate differences in their second Eigenfunction score (Figure 29, Figure 30). Using the linear mixed model can also differentiate phenotypic difference but with lower resolution (Figure 31). All the three SNPs affect CD4 and VL an in opposite way: minor alleles of rs2535307 and kgp22385164 are associated with lower CD4 and higher VL suggesting they are risk alleles, and minor allele of kgp7815029 (chr18, pos45548347, MAF 15%, P_{fpca}=6.22E-06) is associated with higher CD4 and lower VL suggesting it is a protective allele (Figure 28). This result is biologically meaningful.

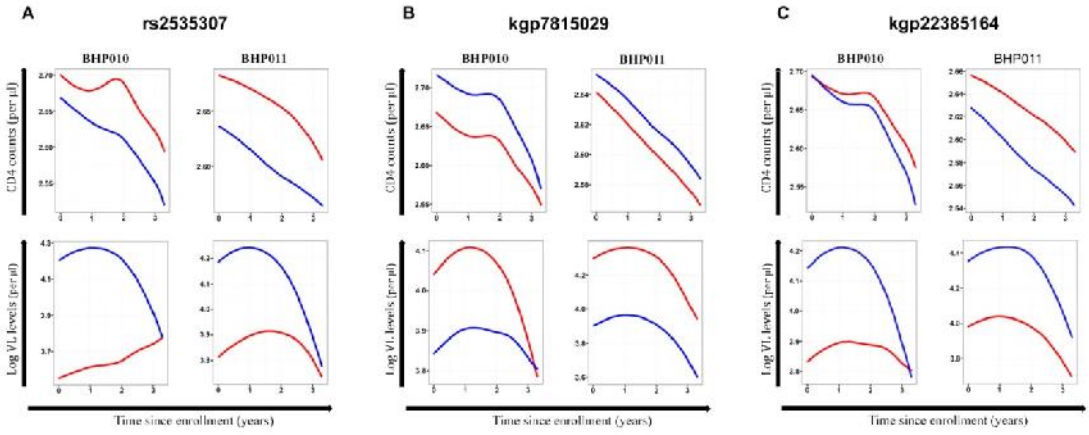

Figure 28. Average trajectories of VL and CD4 are highly corrected with the top significant SNP using dominant model.
Average trajectory of patients with 0 minor alleles (red) or at least 1 minor allele (blue) of significant loci of A.

rs2535307 (chr6, pos31053741, MAF 48%, Pfpca=3.72E-07), B. kgp7815029(chr18, pos45548347, MAF 15%, P_{fpca}=6.22E-06), C. kgp22385164(chr5, pos162859382, MAF=8%, P_{fpca}=9.14E-06). Primary cohort of BHP011 and replication cohort of BHP010 are shown separately. Having at least one copy of minor allele of rs2535307 and kgp22385164 is associated with lower CD4 and higher VL (A and C), reflected by lower value of Eigenfunction 1 of CD4 and higher Eigenfunction 1 of VL compared to patients with zero minor alleles in both BHP010 and BHP011 (Figure 29A,Figure 29C). Minor allele of kgp7815029 is also associated with higher CD4 and lower VL.

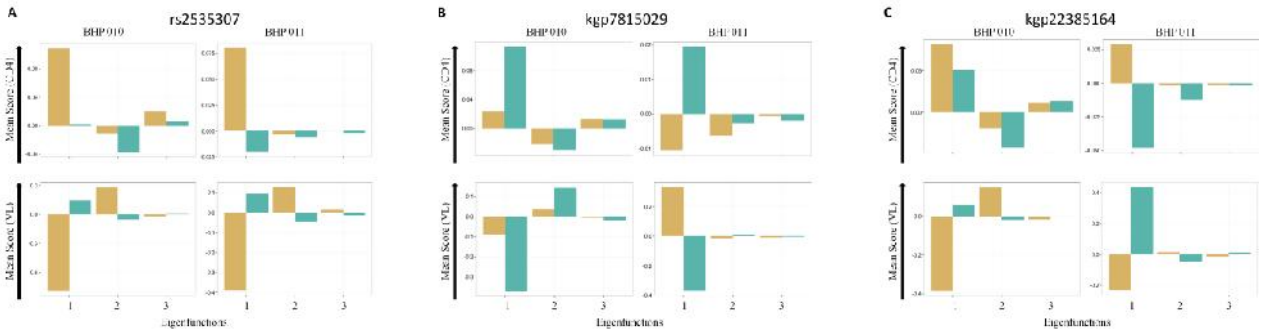

Figure 29. Coefficients measuring the trajectory of VL and CD4 are highly corrected with the top significant SNP using dominant model.

Clinical presentation of patients with alleles of significant loci of A. rs2535307 (chr6, pos31053741, MAF 48%, P_{fpca}=3.72E-07), B. kgp7815029(chr18, pos45548347, MAF 15%, P_{fpca}=6.22E-06), C. kgp22385164 (chr5, pos162859382, MAF=8%, P_{fpca}=9.14E-06). Average fPCA scores Eigenfunction 1 (represents levels of maintaining CD4 or VL), Eigenfunction 2 (represents CD4 or VL decline rate if negative and increase rate if positive), Eigenfunction 3 (variation of the CD4 or VL trajectories). Primary cohort of BHP011 and replication cohort of BHP010 are shown separately. Having at least one copy of minor allele of rs2535307 and kgp22385164 is associated with lower CD4 and higher VL (Figure 28A and Figure 28C). This is also reflected by lower value of Eigenfunction 1 of CD4 and higher Eigenfunction 1 of VL compared to patients with zero minor allele in both BHP010 and BHP011. Minor Allele of rs2535307 and kgp22385164 is also associated with faster CD4 decline and smaller VL changing rate reflected by a more negative Eigenfunction 2 of CD4 and the VL Eigenfunction 2 is closer to zero compared to patients with zero minor allele in both BHP010 and BHP011. The magnitude of effect on Eigenfunction 1 is much larger than Eigenfunction 2, indicating the variants of rs2535307 and kgp22385164 affect the level of CD4 and VL more than changing rate. Similarly, minor allele of kgp7815029 is associated with higher CD4 level and lower VL level indicating their protective role.

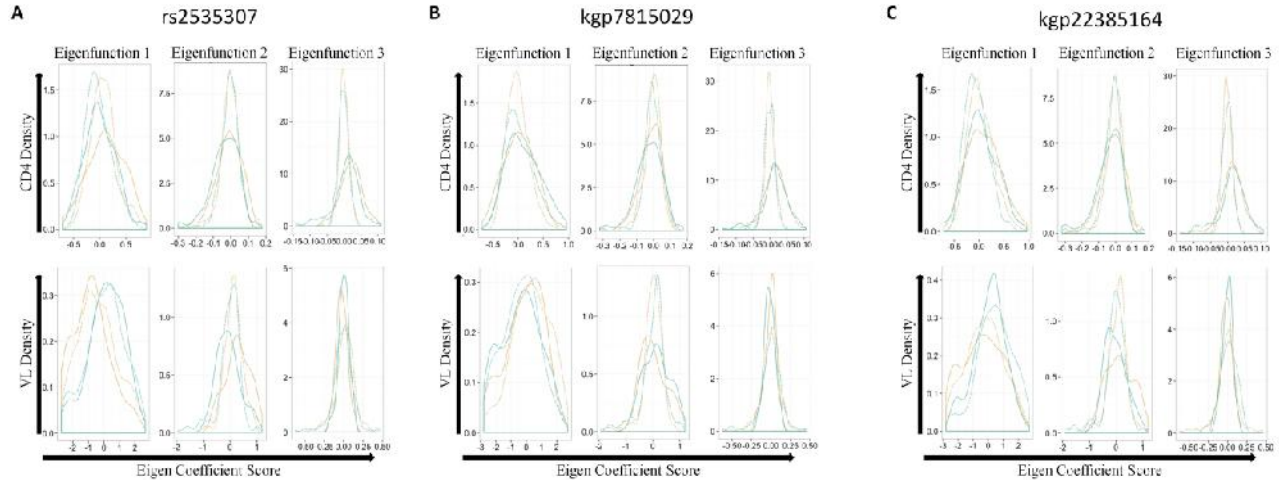

Figure 30. Coefficients measuring the trajectory of VL and CD4 are highly corrected with the top significant SNP using dominant model.

A. The first coefficient of Eigenfunction measuring the average level of CD4 overtime. B. The second coefficient of Eigenfunction measuring the declining rate of CD4 over time. C. The first coefficient of Eigenfunction measuring the average level of VL overtime. B. The second coefficient of Eigenfunction measuring the increase rate of VL over time. Green: Patients with at least one minor allele of "G". Red: Patients with no minor allele. Solid: BHP010. Dash: BHP011.

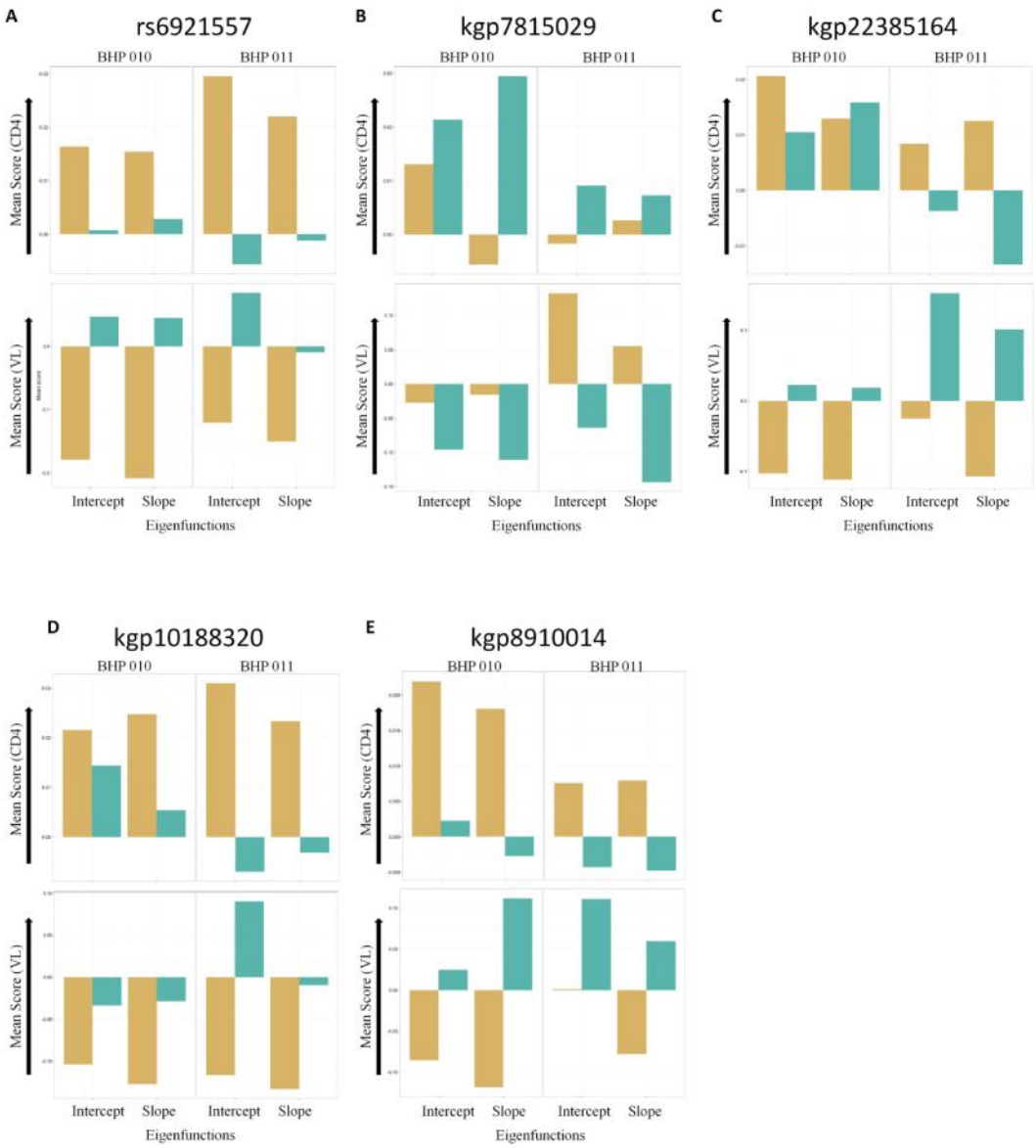

Figure 31. Phenotype presentation of patients with different alleles of top SNPs identified by linear mixed.

Clinical presentation of patients with alleles of significant loci of A. rs6921557 (chr6, pos31049022, MAF 48%, P_{lin}=5.37E-09) B. kgp7815029(chr18, pos45548347, MAF 15%, P_{fpca}=6.22E-06), C. kgp22385164(chr5, pos162859382, MAF=8%, P_{fpca}=9.14E-06),D. kgp10188320, E. kgp8910014. Top: Average VL scores of Intercept (represents baseline), Slope (represents decline rate if negative and increase rate if positive) of patients with 0 minor allele (yellow) or at least 1 minor allele (green); Bottom: Average VL scores. Primary cohort of BHP011 (right) and replication cohort of BHP010 (left) are shown separately.

III.5 Discussion

In this study, we applied both functional principal component analysis and the traditional linear mixed model to define AIDS disease trajectory using longitudinal CD4 and VL. We discovered several new SNPs significantly associated with AIDS disease progression in HIV-Subtype-C infected Botswana treatment naive patients. Using functional principal component analysis, we were able to gain power and identify SNPs that may not be identified using the traditional linear mixed model alone.

a. HCG22 involved in HIV restriction in African population but not others

In both the linear mixed model as well as the fPCA, HCG22 is significant. Previously the SNP rs9262632 in the intron region of HCG22 has been shown to be associated with HIV disease progression in the African American population(International et al., 2010). Our study suggested significant loci of rs2535307(chr6, pos31053741, MAF 48%, P_{fpca}=3.72E-07) and rs6921557 (chr6, pos31049022, MAF 48%, P_{lin}=5.37E-09)near HCG22 which have not been shown before. Patients without the minor allele displayed a significantly higher level of CD4 and lower level of VL (Figure 28A, Figure 31A). It seems the SNP is more related to the level of CD4 and VL more than a decline/increase rate over time. We found that our peak SNP near HCG22 is in high LD with rs9262632 and also the genes within 200kb (Figure 24C) including PSORS1C1, CDSN, CCHCR1, TCF19, POU5F1, PSORS1C3, MUC22 and MUC21. Despite previous GWAS studies identifying HLA-B, HLA-C and HCP5 as strongly associated with AIDS progression in European men(Fellay et al., 2007; Dalmasso et al., 2008; Fellay et al., 2009; Limou et al., 2009; International et al., 2010), it is likely that HCG22 is more significantly related to AIDS progression in African populations. Additionally, the gene cluster of HCG22 has

been reported to be associated total white blood cell counts (Nalls et al., 2011), susceptibility to follicular lymphoma (Skibola et al., 2009) and multiple myeloma (Chubb et al., 2013) in European individuals, suggesting its role in immune cell cycle regulation. This further supported our finding that HCG22 affects CD4 level more than VL. Another GWAS study showed that HCG9, a gene located approximately 1MB upstream of HCG22 on chromosome 6, was a proxy for HLA- A*11:01(Tang et al., 2012). Exon sequencing is needed in the surrounding regions to further identify the causal SNPs for AIDS disease progression in African populations.

b. P21/WAF1 pathway may be highly involved in HIV host restriction in the African population

Interestingly, the other two of our top hits were both on the p53 and p21 pathway that affects cell cycle. Kgp7815029(chr18, pos45548347, MAF 15%, P_{fpca}=6.22E-06) is 5kb downstream of the ZBTB7C gene and is in LD with the intron and coding region of ZBTB7C (Figure 24D). Patients without a minor allele displayed lower CD4 and higher VL compared to patients with at least one minor allele (Figure 28B, Figure 31B). ZBTB7C (APM-1, KR-POK) encodes a zinc finger transcriptional factor that was first described as cotranscribed with HPV DNA in the cervical carcinoma cell line ME180 (Reuter et al., 1998). It has been reported to interact with p53 and regulate the transcription of CDKN1A and p21WAF1, both of which control for cell proliferation(Jeon et al., 2012a). Both p53(Duan et al., 1994) and p21 WAF1 have been shown to be regulated by *tat*, *vpr* and SP1 collaboratively (Sawaya et al., 1998; Amini et al., 2004b) and affect cell cycle (Cherrier et al., 2009). ZBTB7C also interacts with SP1

directly (Jeon et al., 2012b), a transcriptional regulator that directly regulates HIV transcription via interaction with *tat* (Kamine et al., 1991; Sune and Garcia-Blanco, 1995; Zhang et al., 2000; Yedavalli et al., 2003; Gee et al., 2007), *vpr* (Wang et al., 1995; Deshmane et al., 2009) and gag (Zhang et al., 2000). Kgp22385164 (chr5, pos162859382, MAF=8%, P_{fpca}=9.14E-06) is 5kb upstream of the CCNG1 gene and in LD with the coding region of CCNG1 (Figure 24E). Patients without the minor allele had lower VL and higher CD4 (Figure 28C, Figure 29C, Figure 31C). CCNG1 is a member of cyclin family and shares 53% sequence identity with its closest family member CCNG2, which was shown to promote HIV replication by shRNA screen in Jurkat cells (Yeung et al., 2009). Despite CCNG2, multiple other cyclin family members such as CDKN1A (Ryu et al., 2004; Kapasi et al., 2006; Thakur et al., 2012), CDK7 (Nekhai et al., 2002), CCNH (Kino et al., 2002), CCND1(Nelson et al., 2002) have been reported to play roles in HIV replication. HIV *vpr* protein has been shown to arrest the cell cycle by interacting with the cyclins and affecting the function of p53/p21 pathway (Chowdhury et al., 2003; Amini et al., 2004a). It is likely that variation in SNPs close to ZBTB7C and CCNG1 affect CD4 and VL trajectories of AIDS patients via regulation of HIV transcription and the cell cycle of infected cells through the SP1, p53 and p21 pathway. Our association study further raised the possibility of an important role of multiple cross talk between HIV proteins, the p53/p21 pathway and cyclins to regulate HIV replication as well as cell cycle *in vivo* reflected by both viral load and CD4 counts.

c. Multiple genes below significance level may still be associated with host control of HIV/AIDS progression

We also detected several SNPs with combined P-values below a significant level that consistently have small P-values in both cohorts ($P_{combined}<4.4\times10^{-5}$, with both cohort's P value <0.05) (**Table 11**). PRKAA2 can phosphorylate HIV *tat* (Zhang and Wu, 2009; Zhang et al., 2011), *vif* (Shirakawa et al., 2008) and *vpr* (Barnitz et al., 2010; Francis et al., 2011) and regulate their activity directly. ARL8B is a member of the ADP-ribosylation factor family of GTP-binding proteins in which ARL1(Wonderlich et al., 2011) and ARL6 (Blagoveshchenskaya et al., 2002) have been reported to collaborate with HIV-*nef* to disrupt MHC-1 surface level in HIV infected T cells. ARHGEF12is a Rho GTPase which has also been shown to interact with HIV pol in both Jurkat and 293T cells (Jager et al., 2012). ALK is a tyrosine kinase protein that has been reported to interact with HIV-*nef* and played roles in HIV associated dementia (Bergonzini et al., 2009). TRIM family proteins have been reported to be highly involved in innate immunity modulation (Ozato et al., 2008) and anti-retroviral defenses (Nisole et al., 2005) and in our list there are two TRIM proteins that were identified: TRIM26 and TRIM16. TRIM26 is located in the MHC region on chromosome 6 and has been suggested to have roles in inhibiting HIV release. Importantly, TRIM26 was shown as an anti-HIV restriction factor both *in vitro* and *in vivo* such that expression level of TRIM26 has been shown to be elevated in HIV elite controllers compared to HAART suppressed individuals (Abdel-Mohsen et al., 2013) and expression of TRIM26 is correlated with decreased HIV infectivity in PBMC (Raposo et al., 2013).

Table 11 The promising SNPs below genomic significant level

SNP	locus	Annotation	Nearest Gene	Model	MAF (primary)	P-value (replicate)	P-value (primary)	P-value(combined)
rs4912408	1p32.2	intron	PRKAA2	fPCA	0.209375	0.00583	0.000109	9.73E-06
kgp14514681	2q22.1	100kb upstream	HNMT	fPCA	0.095313	2.34E-05	0.029334	1.04E-05

Table 11 The promising SNPs below genomic significant level (Continued)

rs680168	10q26.2	20kb upstream	CLRN3	Linear	0.420313	0.000607	0.001146	1.17E-05
rs10892573	11q23.3	intron	ARHGEF12	fPCA	0.141509	0.003182	0.000303	1.43E-05
kgp10188320	6p22.1	intron	TRIM26	Linear	0.440625	0.065684	1.74E-05	1.45E-05
kgp11221083	16q23.2	intergenic		Linear	0.29375	0.000302	0.003641	1.62E-05
kgp8910014	3p26.1	50kb upstream	ARL8B	Linear	0.1125	0.003224	0.000173	1.99E-05
rs1563011	17p12	intron	TRIM16	fPCA/ Linear	0.121875	0.000228	0.006107	2.02E-05
kgp11815484	2p23.2	intron	ALK	fPCA/ Linear	0.260252	3.98E-05	0.047328	2.67E-05
kgp11537425	13q21.2	intergenic		Linear	0.329688	0.009521	0.000204	2.75E-05
rs2504890	6p24.1	intergenic		Linear	0.423438	0.01238	0.00014	2.81E-05
kgp18828163	12p13.32	intron	AK095365	fPCA	0.126563	0.037386	5.62E-05	2.96E-05
rs12111444	6p12.1	100kb upstream	KLHL31	fPCA	0.289969	0.001845	0.001318	3.39E-05
kgp22829958	Xq21.32	intron	PCDH11X	fPCA	0.184375	0.018558	0.000167	4.24E-05
kgp3432145	4p16.3	exon	ZFYVE28	fPCA	0.063091	0.006051	0.000525	4.34E-05

d. PCDH11X on X chromosome may be associated with HIV restriction in vivo

We excluded X and Y chromosomes in our association analysis to avoid potential bias if adjusting for sex and age. We also replicated the study including the sex chromosomes on 852889 SNPs and before adjusting for age and sex, we found an additional interesting SNP lying in the intron of PCDH11X (peak variant kgp22829958, Pcombined=4.24×10^{-5}) (Table 11). PCDH11X is a member of protocadherin family that knock down PCDH11X via siRNA and has been shown to affect HIV transcription in TZM-bl cells (Brass et al., 2008). After adjusting for age and sex, we got an additional 3 new SNPs with low P-

value (错误！未找到引用源。). The top SNPs before adjustment were unaffected by adjustment (Figure 32C).

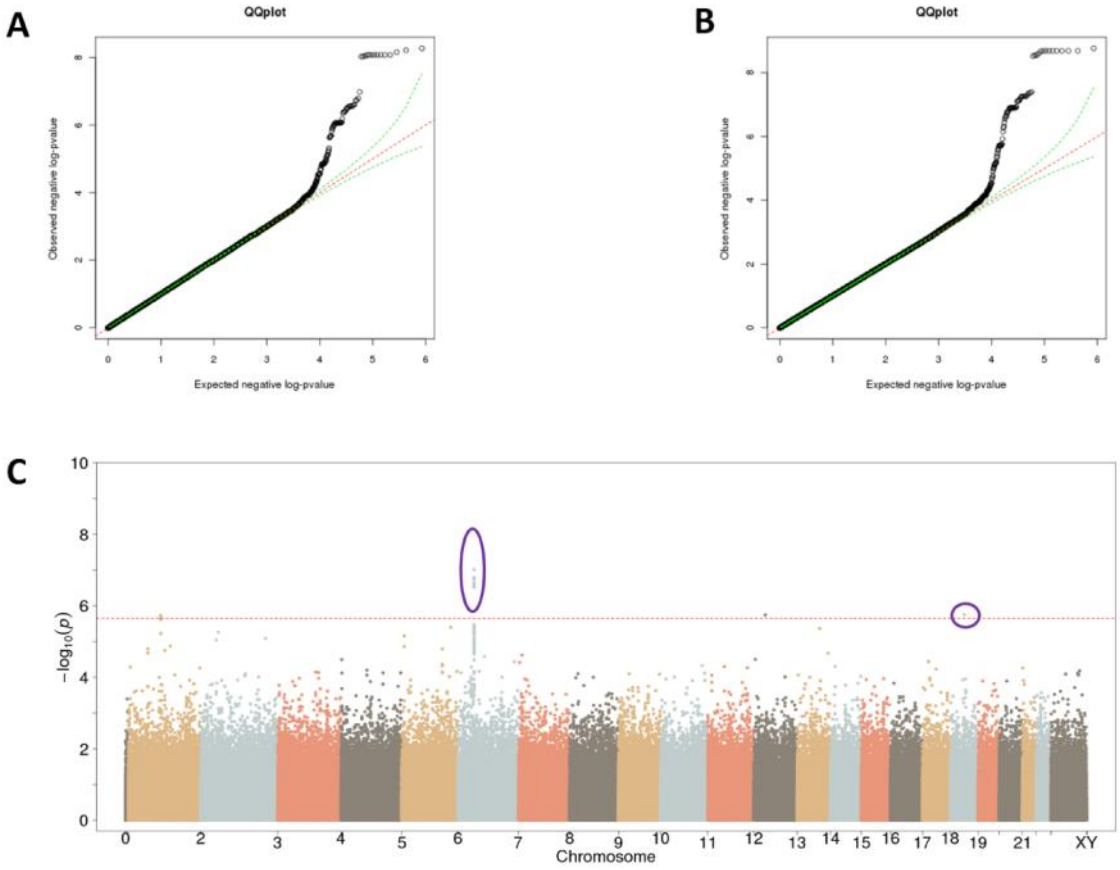

Figure 32. The QQ plot and Manhattan plot after adjusted by age and sex if including sex chromosome in association analysis (dominant model) for fPCA.
A. QQ plot using fPCA model after adjusting for age and sex, including X and Y chromosomes. B. QQ plot using linear model after adjusting for age and sex, including X and Y chromosomes. C. Manhattan plot using fPCA model after adjusting for age and sex, including X and Y chromosomes. Top SNPs identified before stratification still remain top. Four independent SNPs passed significant thresholds and HCG22 is the only one with a consistent small p-value in both cohorts. All sex chromosome SNPs failed to reach the genome-wide significant level after gender and age adjustment.

e. New SNPs with small p-value after adjusting for age and sex

Previous studies of GWAS for both the "Centre for HIV/AIDS Vaccine Immunology study "(CHAVI) Study (Fellay J, 2007) and the microsatellite-based GWAS of SIV infected monkeys showed that an unknown X chromosome variation is associated with slow AIDS progression defined by a lower viral load at set point and slower CD4 decline.

Strikingly, this effect is only in HIV infected females but not males (Siddiqui RA, 2009). Aging is also associated with alterations in T-cell function and decreasing ability to reconstitute CD4 T-cells in HIV patients after antiviral treatment (Rickabaugh TM, 2010; Appay V, 2011). These pieces of evidence suggest gender and age are potential effect modifiers for host genes associations with CD4 trajectory.

We also found three interesting SNPs with small p-values after adjusting for age and sex, indicating these SNPs may affect HIV host control in specific gender or age groups (**错误！未找到引用源。**). The genes involved are mir4500HG (peak variant kgp16854268,Pcombined=4.35×10^{-6}), CHIAP2 (peak variant rs3122595, Pcombined=1.77×10^{-5}) and FHIT (peak variant kgp5164069, Pcombined=2.02×10^{-5}). mir4500HG is a ncRNA that hasn't been linked to a particular function. CHIAP2 is a pseudogene of chitinase and also has no report of function. Chitinases have been reported to have important roles in plants and insects (Nagpure et al., 2013). Humans have 8 Chitinases (GH18) but most bear essential mutations on a glutamic acid residue at the catalytic domain so the protein is not functional (Guan et al., 2009). FHIT (fragile histidine triad) encodes a diadenosine 5',5'''-P1,P3-triphosphate hydrolase involved in purine metabolism. It has been identified as a tumor suppressor gene in multiple cancer types (Wu et al., 2003; Cecener et al., 2010; Al-Temaimi et al., 2013). Yu GR et.al used a HIV-*tat* fused FHIT to deliver the pro-apoptotic protein to hepatocytes and showed inhibition of hepatocellular carcinoma(Yu et al., 2012). However there hasn't been any report of FHIT's role in HIV restriction.

Table 12 New SNPs with small p-values after adjusting for age and sex (using fPCA, dominant model)

SNP	locus	Annotation	Nearest Gene	Model	MAF (primary)	P-value (replicate)	P-value (primary)	P-value(combined)
kgp168 54268	13q31.2	intron	mir4500HG	fPCA/ Linear	0.0721	1.73E-05	0.01559	4.35E-06
rs31225 95	1p13.2	10kb upstream	CHIAP2	fPCA/ Linear	0.453125	9.86E-05	0.012285	1.77E-05
kgp516 4069	3p14.2	50kb downstream	FHIT	Linear	0.143145	0.004216	0.000332	2.02E-05

f. Using fPCA allows better simulation of small variations in longitudinal AIDS clinical data and identifies new SNPs

We noticed that except for HCG22, all other novel SNPs we discovered haven't been reported to be associated with AIDS disease phenotypes in earlier reports. We thought this may be due to two reasons. Firstly, we applied a new mathematical method (fPCA) which can better adapt the variation of trajectory. Previous GWAS studies applied linear mixed method to obtain individual viral loads or CD4 count slopes which assumed linear trajectory shape and usually excluded certain CD4 and VL counts manually to best fit for the assumed linear trajectory(Fellay et al., 2007). Linear shape displays the lowest degree of freedom to capture the variation. However the pathogenesis of HIV is complicated in that both CD4 and VL trajectories can be non-linear. Consequently the linear mixed method loses resolution and scores by the linear mixed method are more subject to misclassification of patients' real phenotypes(SE, 2012). Secondly, given that the SNPs reported before were mostly shown by GWAS conducted on Caucasian men infected with HIV-1-subtype B, it is possible that host genome uses some different mechanisms in controlling HIV-1-Subtype-C and the larger variation within the African population enables us to discover new SNPs. We raised this hypothesis because even though we

replicated the previous study using their reported linear mixed method (Fellay et al., 2007) on our cohorts, the p-value of SNPs suggested by previous GWAS that strongly associated with AIDS VL or disease progression (rs9264942, rs9261174 and many other SNPs identified to be associated with AIDS progression or CD4/VL levels) (Fellay et al., 2007; van Manen et al., 2012) were not among the top100000 smallest p-values in our study (Table 13). We failed to examine the top hit in previous study of rs2395029 due to the fact that this SNP's MAF is lower than 0.05 (Table 13) in both of our cohorts.

Table 13 P-value of SNPs suggested before by other GWAS

SNP	Phenotype	Ethnicity	Nearest Gene	Linear model Combined p value	fPCA model Combined p value	Cohort
rs2395029(Fellay et al., 2007; Dalmasso et al., 2008; Fellay et al., 2009; Limou et al., 2009; International et al., 2010) (kgp10902)	RNA VL set point, plasma HIV RNA primary infection, cellular HIV DNA primary infection, long-term non-progression, VL controllers	Caucasian	HCP5	Minor allele frequency <0.05		CHAVI, ANRS PRIMO, GRIV
rs9264942(Fellay et al., 2009; Siddiqui RA, 2009; International et al., 2010)	RNA VL set point, VL controllers		HLA-C	0.856069	0.869542	CHAVI
rs9261174(Fellay et al., 2007)	CD4 T cell decline		ZNRD1	0.076274	0.070016	CHAVI
rs9368699(Limou et al., 2009)	long-term non-progression		C6orf48	Minor allele frequency <0.05		GRIV
rs4118325(Le Clerc et al., 2009)	rapid progression		PRMT6	0.328242	0.288011	GRIV
rs1522232(Le Clerc et al., 2009)	rapid progression		SOX5	0.630198	0.470801	GRIV
rs17762192(Herbeck et al., 2010)	progression to AIDS		PROX1	0.649821[¶]	0.261922[¶]	MACS
rs2234358(Limou et al., 2010)	long-term non-progression		CXCR6	0.441716	0.539889	GRIV, MACS, ACS
rs152363(van Manen et al., 2011)	progression to AIDS, or AIDS-related death		AGR3	0.616375	0.641990	ACS
rs4418214(International et al., 2010)	VL controllers	Caucasian, African, Hispanic	MICA	Minor allele frequency <0.05		International HIV controllers study
rs3131018(International et al., 2010)	VL controllers	Caucasian, African, Hispanic	PSORSIC3	0.817738	0.630018	International HIV controllers study
rs2523608(Pelak et al., 2010)	RNA VL set point	African	HLA-B*5703	0.221724	0.239658	DoD HIV NHS and MACS
rs11884476(Troyer et al., 2011)	progression to AIDS	Caucasian	PARD3B	0.105197[Ψ]	0.153524 [Ψ]	MACS

[¶]The original SNP rs17762192 is not on the Chip. We used the closest SNP kgp15424027which is 103 bp upstream of rs17762192.

[Ψ]The original SNPrs11884476 is not on the Chip. We used the closest SNPkgp1244933 which is 673 bp upstream of rs11884476.

Chapter IV Discussion

IV.1 Using human pluripotent stem cell as an in vitro model of HIV latency

The concept of a latent reservoir for HIV arose in the 1990s from clinical observations of residual cells harboring the HIV proviral genome and the treatment of AIDS patients that could maintain low viral load for a period of time. The natural HIV reservoir was first identified as resting CD4+ T cells(Chun et al., 1997). These cells have a very slow decay rate so that they serve as a major barrier to viral eradication by HAART (Finzi et al., 1999; Zhang et al., 1999a; Ramratnam et al., 2000; Siliciano et al., 2003; Strain et al., 2003; Han Y, 2007b). Besides resting CD4+ T cells, it has been suggested that other cell types can also serve as HIV proviral reservoirs including hematopoietic stem cells (Carter CC, 2010), macrophages (Zhang et al., 1999b; Igarashi et al., 2001), dendritic cells(McIlroy et al., 1995; Donaghy et al., 2001) and monocytes (Lambotte et al., 2000; Sonza et al., 2001; Zhu et al., 2002; Ellery et al., 2007; Jaworowski et al., 2007; Hasegawa et al., 2009). However latently infected cells represent a very low fraction of all blood cells (for CD4+ T cells about 1 integrated copy of HIV in 10000 CD4+ T cells) (Chun et al., 1995). Additionally, the definition of "latently infected cells" states that the infected cells are reversibly infected. Such cells produce little HIV-1 mRNA, and the mRNA produced are usually either prematurely terminated(Brooks et al., 2003; Hermankova et al., 2003) or mislocalized(Kao et al., 1987; Adams et al., 1994; Lassen et al., 2004). Therefore the latent cells have no viral proteins as markers and can only be distinguished from uninfected cells by integrated proviral DNA. The latent reservoirs may get reactivated under certain conditions and produce infectious viral particles.

Because of the importance of HIV reservoirs as an obstacle to curing AIDS, it is crucial to have more *in vitro* models that allow scientists to study the mechanism of HIV latency. The exploration of reactivation pathways for drug screening may lead to new strategies for treatment. The most ideal *in vitro* model of HIV latency might be cells from patients that harbor latent HIV genomes, especially if the cells were available for culturing and proliferation *in vitro* to perform intervention experiments. However, due to the fact that we still don't have an effective tool to survey the proviral HIV integrated at a single cell level without killing the cells, that the frequency of latent cells are extremely low *in vivo*, and that latent cells cannot be effectively propagated *in vitro*, new approaches are needed to provide systems for analysis.

Surrogate cell models, such as Jurkat cells, Hela cells and 293T cells, have been widely used to study HIV latency. In fact, many important discoveries of HIV latency (mostly at epigenetic level)were made using those cell types(reviewed in II.2). However it has been argued that the active machinery in transformed cell types are different from untransformed cells(Chen and Dent, 2014). Therefore the relative value of findings derived from the *in vitro* cancer cell models remains to be decided.

Human pluripotent stem cells are not a natural host for HIV due to the lack of CD4 HIV entry receptors and the co-receptors. There has been no evidence showing HIV is able to infect embryonic stem cells during natural infection. It has also been shown that sperm are also not targets of HIV infection (Kim et al., 1999), though washing sperm has been a standard practice to decrease the chance of having HIV positive babies born to

serodiscordant-couples(Pasquier et al., 2000; Garrido et al., 2006). Therefore our rationale of using human embryonic stem cells was not to mimic natural infection.

What makes human pluripotent stem cells appealing as a new model for HIV latency are the following characteristics:

i). Theoretically human pluripotent stem cells are able to differentiate into all the lineages of adult cells (Murry and Keller, 2008; Hanna et al., 2010; Yabut and Bernstein, 2011), including all the natural host cell types of that can be infected by HIV. Due to the limited knowledge of developmental biology and cell differentiation techniques, we cannot derive all the host cell types that HIV infects *in vitro*, but the current cell types that we can derive from human pluripotent stem cells include hematopoietic progenitor cells (Vodyanik et al., 2005; Doulatov et al., 2013), T cells(Galic et al., 2006), macrophages (Senju et al., 2011; van Wilgenburg et al., 2013) and dendritic cells(Li et al., 2014). Potentially we will be able to derive some natural reservoirs of HIV using the iPS cell model we generated that harbors latent HIV. Such a model could perhaps also be used for drug screening *in vitro*.

ii). The pluripotent stem cells' epigenetic signature is similar to that of hematopoietic stem cells(Spivakov and Fisher, 2007). The epigenetic signature on the genome of human pluripotent stem cells displays a bivalent profile that many markers such as H3K9ac, H3K4me and H3K27me tend to be more enriched (Efroni et al., 2008; Cui et al., 2009). Many essential HIV transcription activators or repressors such as SUV39H1(du Chene et al., 2007) and HDACs(Zhang et al., 2011) are also actively participating in establishing

the chromatin signature of pluripotent stem cells or hematopoietic stem cells. Due to the fact that hematopoietic stem cells cannot effectively be cultured ex vivo, the embryonic stem cells used represent the best approximation to hematopoietic stem cells.

To better adapt the model to studying patient-specific mechanisms for HIV latency, we proposed to use patient cell-derived iPS cells generated from PBMC that harbor HIV viruses. Based on our study, we found that the embryonic stem cells were able to silence the full-length HIV virus. After reprogramming the blood cells from patients into the iPS stage, we can further re-differentiate the patient-derived iPS towards blood cell lines again (Figure 33). Those cells will harbor HIV proviral genes, and patient specific genetic mutations. According to our finding of long-term HIV latency, those proviruses seem likely to be silenced in the re-differentiated blood cells. By doing so we will be able to generate patient-specific HIV reservoirs that harbor latent HIV. Importantly this strategy may allow us to expand patient-specific iPS cells *in vitro* infinitely and derive the HIV blood cell reservoirs for research and drug screen usage (Figure 33). Theoretically, we will be able to perform personalized drug screens on molecules that will potentially be able to reactivate the HIV latent viruses on these patient-derived HIV reservoirs. However, many technical difficulties may still need to be overcome with this strategy and more research is necessary.

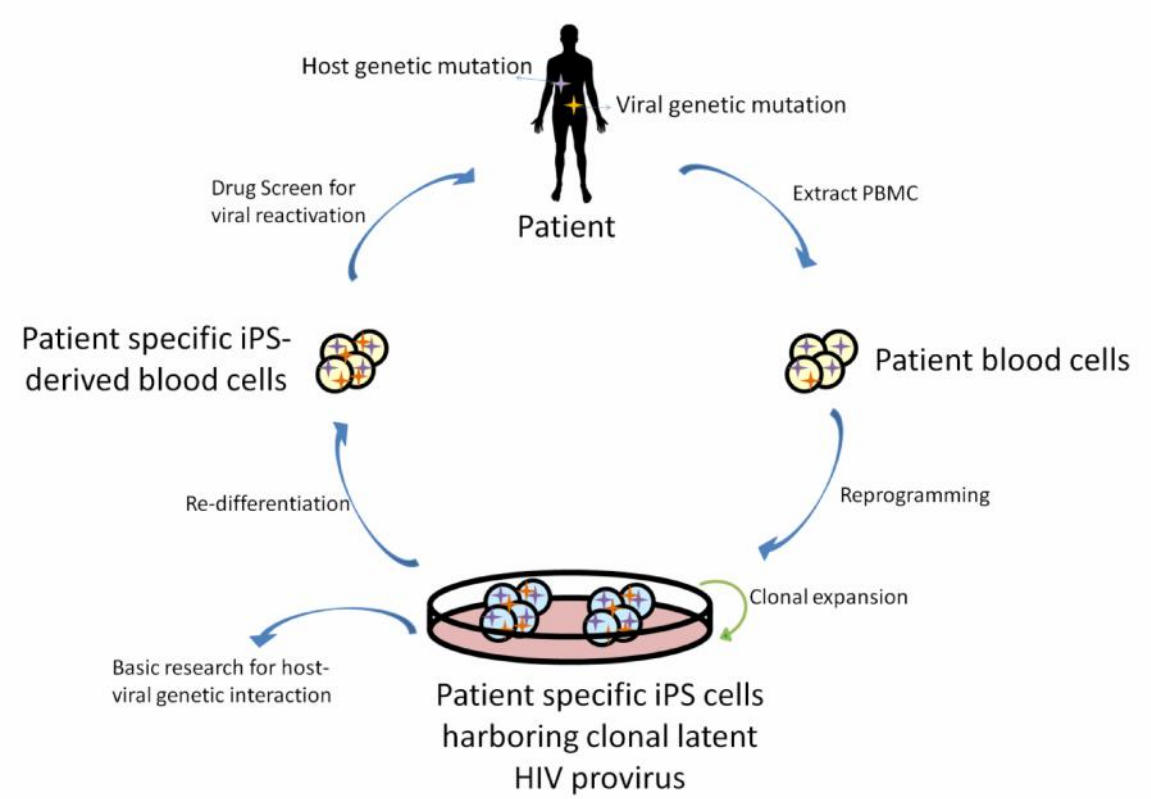

Figure 33. Strategy of using iPS cells as patient specific HIV model for both drug screen and basic research

IV.2 Implication of host restriction to HIV replication by transcription factors

We suggested that ZBTB7C and CCNG1 are associated with HIV disease progression in treatment naïve African patients infected with HIV-1C. As discussed in III.5b, these two genes are both in the p53 and p21 pathway that affect cell cycle. Though there hasn't been direct reports showing these two genes are transcriptional factors, they have been shown to interact with p53 and p21, two very important transcriptional factors regulating the cell cycle.

Despite the fact that most of the GWAS on HIV identified hits around HLAs, previous GWAS studies have also identified several transcriptional factors including ZNRD1 (Fellay et al., 2007), PARD3B (Chinn et al., 2010), SOX5 (Le Clerc et al., 2009) and PROX1 (Herbeck et al., 2010). In vitro, many other transcriptional factor pathways including the NF-KB pathway (Antoni BA, 1994), Sp1 (Kamine et al., 1991), activator protein 1 (AP-1) proteins (Gonzalez et al., 2001), nuclear factor of activated T cells (NFAT) (Lahti et al., 2003), and CCAAT enhancer binding protein (C/EBP) (Dahiya et al., 2014) were identified to be key regulators of HIV transcription. However in this *in vivo* association study we only identified two genes involved in transcriptional regulation, both in the same p21/p53 pathway. Our result raises the importance of p21/p53 pathway in regulating HIV transcription *in vivo*. However, other pathways that were shown to be important *in vitro* were not among the top hits *in vivo*, despite the sophisticated cross-talk among these pathways. We think this may be due to two reasons: First, many of these genes may be crucial for cell function so that variation in the genes that affect protein functions will be very rare in the population. Second, HIV genome transcription is governed by multiple transcriptional pathways as discussed above. Because of the cross-talk among these pathways and the redundancy of the players, the effect of mutations in one pathway may be rescued by another so that the effect of transcriptional factor variation may be subtle and hard to detect at the level of CD4 and VL. Our novel application of FPCA is able to capture minor variations in CD4/VL trajectories that the linear mixed model cannot, yet some effects may still be too subtle for detection and more parameters need to be measured in order to better define the phenotype. We,

however, haven't looked at proviral load that may be one parameter we can measure to study the effect of HIV host control at the transcriptional level.

Additionally, though there has been a report showing that ZBTB7C is cotranscribed with HPV DNA in the cervical carcinoma cell line ME180(Reuter et al., 1998) and it may interact with p53 to regulate p21WAF1 which in turn interacts with multiple HIV accessory proteins (reviewed in III.5b), there hasn't been a report on ZBTB7C and HIV replication *in vitro*. Similarly, for CCNG1, though its close family member CCNG2 was shown to promote HIV replication by shRNA screen in Jurkat cells, there has been no report of CCNG1 function in HIV transcriptional control. It will be interesting to check these two proteins *in vitro* to test their functionality in HIV transcription restriction.

In II.4j, using *in vitro* experiments, we identified that the HIV accessory protein Nef serves upstream of PAK2 to act as a transcriptional regulator. HIV Nef, activated by myristoylation by NMT1 and NMT2, activates PAK2 and other host factors such as Hck. Full length PAK2 is autophosphorylated when activated by p21(Pirruccello et al., 2006) and Nef. It is possible that downstream transcriptional factors regulated by PAK2 interact with the HIV genome and modulate HIV transcription in human pluripotent stem cells (Figure 34).

Combining our *in vivo* and *in vitro* studies, we raised the following model in Figure 34. Our *in vivo* study raises the possibility that host factor of ZBTB7C may act upstream of p53 and p21 to inhibit multiple cyclins including CCNG1, to arrest the cell cycle. This

could possibly affect CD4 counts. The activation or inhibition of p53 is regulated by HIV accessory proteins *tat* and *vpr* (Figure 34). Furthermore, ZBTB7C interacts directly with SP1 to activate the HIV transcription, which will directly affect viral load. CCNG1 interacts with Vpr to repress HIV transcription. The opposite function of ZBTB7C and CCNG1 is also consistent with our *in vivo* observation that patients having minor variation of ZBTB7C or CCNG1 have opposite phenotypes regarding CD4 and VL levels.

Overall both our *in vitro* and *in vivo* studies suggest that the p21 pathway played a central role in HIV pathogenesis.

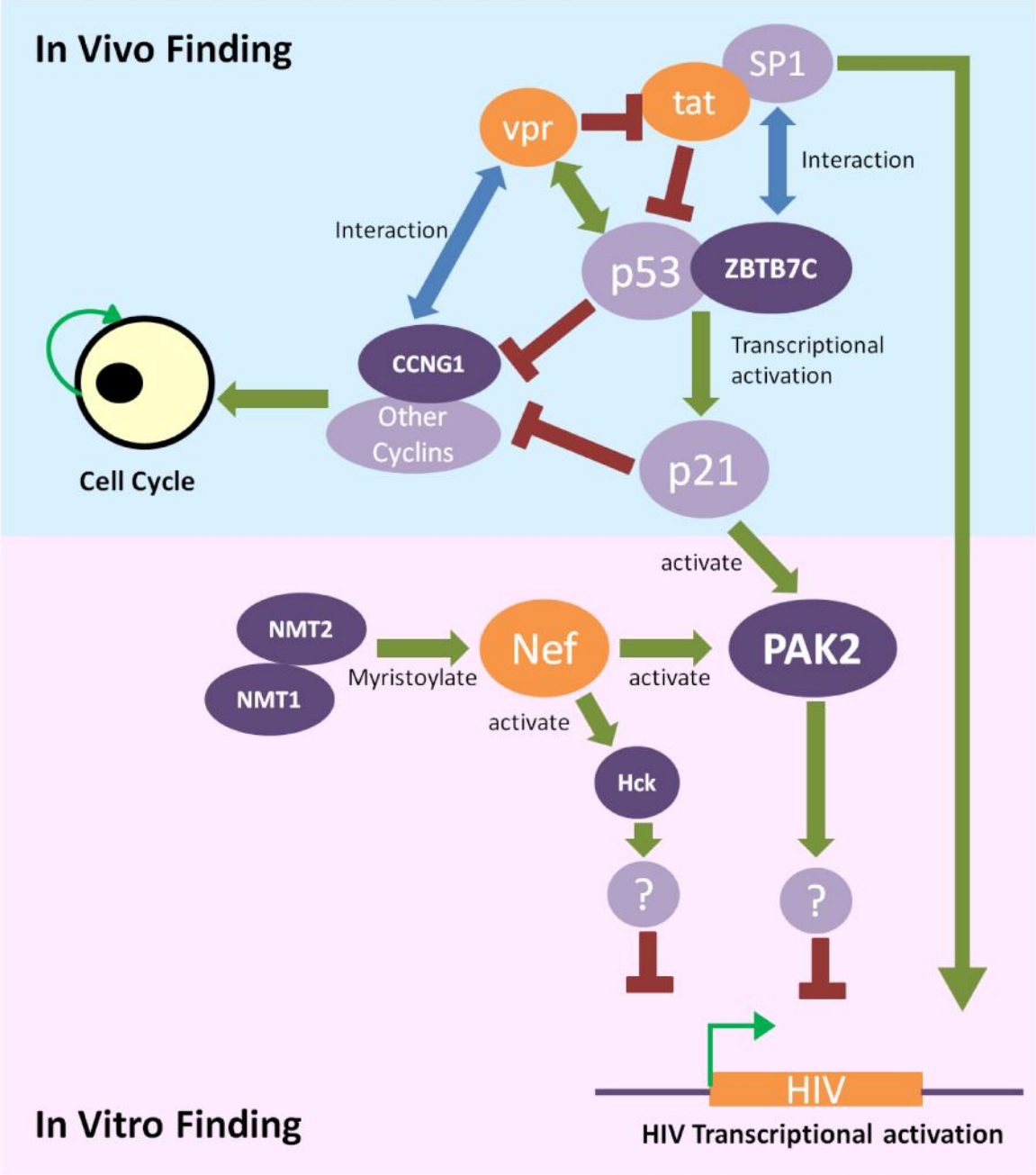

Figure 34. In vitro and In vivo combined finding of HIV viral factors collaborating with host factors to regulate HIV replication and host cell cycle.
Purple: Host factors (Dark Purple: factors identified in the study; Light Purple: factors suggested by literature); Orange: Viral factors; Green arrows: activating function; Red: inhibiting function; Blue: interaction that hasn't been characterized. Blue background: pathways identified by *in vivo* GWAS study; Pink background: pathways identified by using pluripotent stem cells *in vitro*.

IV.3 The collaboration of host genes and viral genes in controlling HIV replication and affecting CD4 and viral RNA

Our research showed that three genes are associated with CD4/VL trajectories over time as a combined phenotype. From an epidemiology perspective, HIV infection and progression to AIDS is a complicated process where host genes, viral properties and environmental exposures are all involved in determining CD4 and VL trajectories (Figure 35A). Previous GWAS showed that the variation within HLAs, ZNRD1 and RFN39 can only explain less than 10% of the phenotypic variation in outcomes (Fellay et al., 2007). Environmental factors such as previous exposure to drugs, nutrition, education, and viral genome variation may also affect the phenotypes. Host genes, viruses and environmental factors serve as the causal factors, interacting with each other in the human body. By using the methodology of GWAS, we can only infer a fraction of the complicated interaction among these causal factors by establishing an association between genetic alleles and measurable outcomes such as CD4 and VL (Figure 35A).

Host factors exert their role of viral restriction by interacting with viral elements, while viral factors facilitate viral replication by interacting with host genes. Viral and host genes are always interacting with each other to play roles in viral replication restriction. Identifying the viral target of host genes and stage of viral restriction in the HIV life cycle is very important if we want to design HIV interventions.

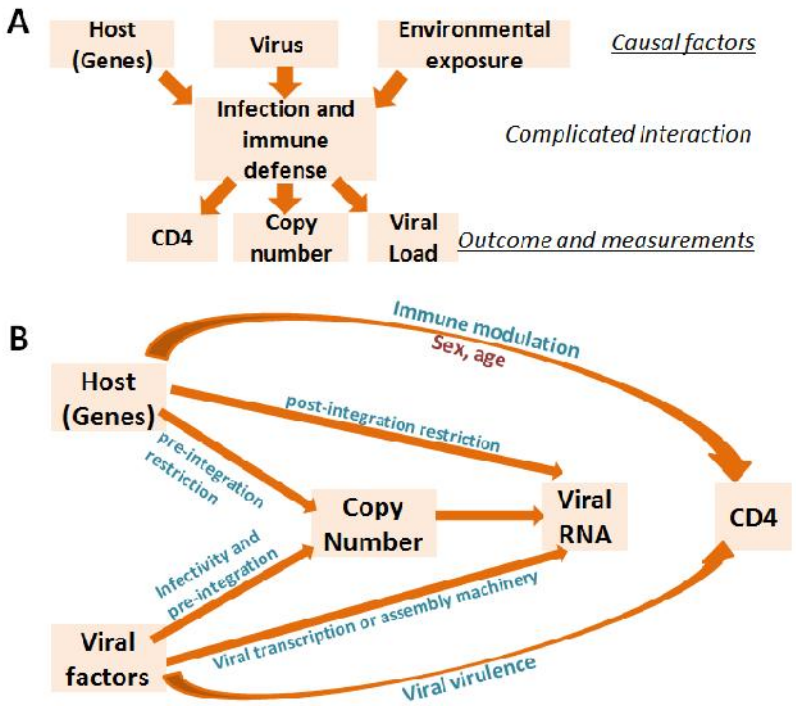

Figure 35. Epidemiology modeling of HIV-1 infection.
A. Models depicting factors that could affect the read-outs of HIV infection. B. causal DAG model showing the causal relationship of factors/exposure and the GWAS study read out. Blue: The candidate mechanism of casual inference; Red: candidate effect modifiers to the causal inference.

Before virus gets into the cells, host cell surface receptors CD4 and CCR5/CXCR4 interact with gp120 to facilitate entry of HIV into host cells. Then host genes such as APOBEC3G and LEDGF/p75 interact with *vif* and integrase respectively to control viral replication before integration (Anderson et al., 2009). To isolate the host viral interaction before integration, GWAS studies based on populations with high exposure to HIV infection (such as sex workers) use HIV positivity as phenotype and tested the association between host genes and HIV susceptibility (An et al., 2014). After integration, multiple host genes are involved in viral transcription, or regulating HIV transcription indirectly by chromatin modification. It is very hard to isolate the host genetic restriction mechanisms at a transcriptional level by GWAS. Here we suggest that by surveying HIV proviral copy number, it may be possible to detect extreme phenotypes

of patients with heavy proviral loads but low viral RNA or the opposite (Figure 35B). Though previous studies have shown some positive relationship between HIV proviral load and viral load, the association is not strong (Novitsky et al., 2006) (Figure 36). We suggest this new defined phenotype combining proviral load and viral load may help capture the subtle effect of transcription factors and other post integration restriction mechanisms *in vivo*.

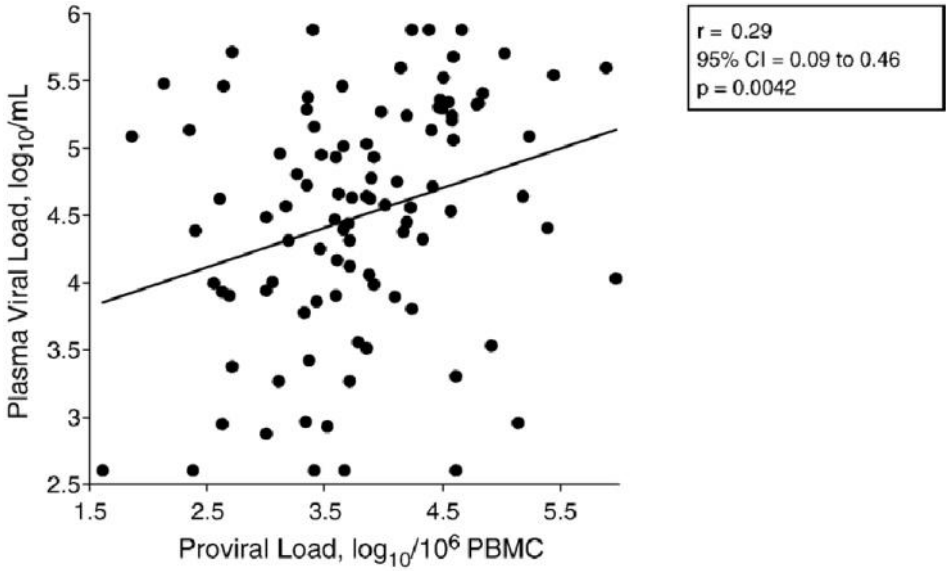

Figure 36 Association between proviral load and plasma viral load in HIV-1C infection
From reference (Novitsky et al., 2006)

The CD4 level of patients can be affected by factors other than those that directly affect replication. Immune regulation factors such as multiple interleukins that affect cell proliferation, and factors associated with control of cell death, may also be possibly involved in affecting the CD4 level of patients. On the other hand, viral factors that determine viral virulence can interact with host cell death pathways to influence the cell cycle. Host and viral factors that affect cell cycles should have stronger associations with CD4 level as compared to VL (Figure 35B). There is another interesting observation in

our study, that if we use only CD4 phenotype, though the top few SNPs are still similar, their p values are much larger except for region around rs2535307(chr6, pos31053741, MAF 48%, P_{fpca}=3.72E-07) (data not shown). This observation suggested that the HCG22 region is more associated with CD4 counts while ZBTB7C or CCNG1 may be more associated with VL. However, since we haven't tested the role of ZBTB7C and CCNG1 in viral restriction *in vitro*, and since these two genes are highly involved in cell cycle pathways (Figure 34), it is also possible that ZBTB7C and CCNG1 are affecting both cell cycle and viral load . To isolate the host genes' effect in regulating cell proliferation, an analysis to examine the association between host genes and CD4 or VL alone, rather than a combined phenotype of CD4 and VL, may be needed.

IV.4 Using genetic markers as a public health indication for AIDS treatment prioritization

From a public health stand point, treatment prioritization is crucial for improving quality of life and averting premature death in adults and children living with HIV. In Southern Africa, due to the high rate of AIDS and economic constraint, not all patients are able to get HAART upon detection of HIV positivity. Instead, an approach was adopted based on WHO guidelines: in Botswana, only patients with CD4 counts lower than 350 cells/μl on clinical illness are eligible for HAART treatment. Patients with CD4 higher than 350 cells/μl and lower than 500 cells/μl are eligible for periodic follow up for every three months(Government-of-Botswana, 2012). Patients with CD4 higher than 500 cells/μl are followed every six months. However, due to the disease progression variability of AIDS patients, the patients not immediately eligible for HAART may have different rates of

progression to severe disease development. For some patients (Figure 37 upper panel), it took less than half a year for CD4 to decline from above 500 cells/μl to less than 200 cells/μl, while other patients can maintain their CD4 around 500 cells/μl for at least 4 years (Figure 37 lower panel). For patients with low CD4 counts, even short delays (less than one month) in initiating treatment are associated with increased mortality. Therefore eligibility of initiating HAART based on arbitrary CD4 levels is not an optimal approach.

However, it is often difficult to identify patients that are predisposed to abrupt CD4 decline. We established an association between host genes and disease progression using both CD4 and VL as phenotype. We found three alleles that are strongly associated with disease progression. Using the combination of the three alleles, it may be possible to develop algorithms to help predict CD4 trajectory of each patient, and to use this prediction as an indicator for HAART eligibility.

For example, we randomly picked three patients from BHP010 that showed fast decline of CD4 within 6 months, and three patients with stable CD4 trajectories from BHP010 (Figure 37). Their genotypes of the three SNPs, which we identified to be strongly associated with disease progression (Table 10), are listed below their phenotype plots. The patients that are fast-progressors all carry two risk alleles of rs2535307(chr6, pos31053741, MAF 48%, P_{fpca}=3.72E-07), and some carry another risk allele kgp22385164(chr5, pos162859382, MAF=8%, P_{fpca}=9.14E-06), though all of them carry protective alleles at kgp7815029(chr18, pos45548347, MAF 15%, P_{fpca}=6.22E-06). For the three slow-progressors, none of them carry risk alleles rs2535307 and kgp22385164.

It is worth noticing that though this is not a study designed for phenotypic prediction, by random sampling the cohort, it is possible to use the three SNPs to predict the disease progression for these patients. And since these three alleles are significantly associated with the phenotype, it is possible to build statistical simulation models to predict patients' AIDS progression using the three genes identified. It also worth studying if rs2535307 alone is sufficient in predicting the disease phenotype as it has the strongest association in our result. Ideally if the model works well, we can potentially design a genotyping kit to survey these SNPs among patients that are HIV positive but not yet eligible for HAART due to higher CD4 levels. Such a result could help inform HAART prioritization in people carrying risk alleles that are linked to faster AIDS progression.

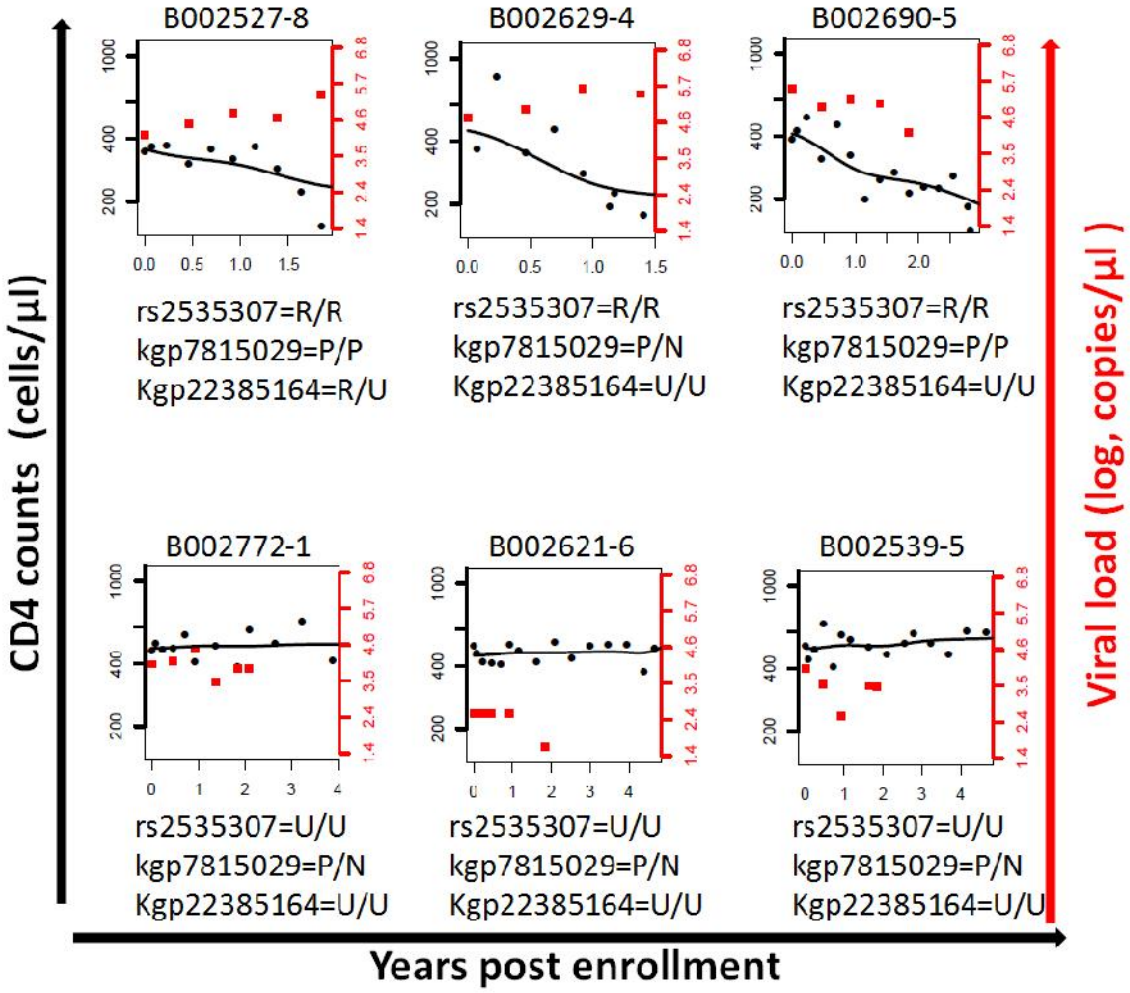

Figure 37. Extreme phenotypes of CD4 decline and genotype of the three significant SNPs.
U: non-risk allele, R: risk allele. P: protective allele, N: non-protective allele. Using dominant model, one risk allele or one protective allele will result in dominant effect. Upper panel: patients that experience fast CD4 decline from above 500 cells/μl to <200 cells/ml within half a year. Lower panel: patients that are able to maintain CD4 around 500 cells/μl for at least four years.

Bibliography

(2012). UNAIDS report on the global AIDS epidemic (WHO Library Cataloguing-in-Publication Data).

(2012). AIDSinfo: epidemiological status. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS) ((http://www.unaids.org/en/dataanalysis/datatools/aidsinfo/)).

Abdel-Mohsen, M., Raposo, R.A., Deng, X., Li, M., Liegler, T., Sinclair, E., Salama, M.S., Ghanem Hel, D., Hoh, R., Wong, J.K., *et al*. (2013). Expression profile of host restriction factors in HIV-1 elite controllers. Retrovirology *10*, 106.

Adams, M., Sharmeen, L., Kimpton, J., Romeo, J.M., Garcia, J.V., Peterlin, B.M., Groudine, M., and Emerman, M. (1994). Cellular latency in human immunodeficiency virus-infected individuals with high CD4 levels can be detected by the presence of promoter-proximal transcripts. Proceedings of the National Academy of Sciences of the United States of America *91*, 3862-3866.

Ahmad, N., and Venkatesan, S. (1988). Nef protein of HIV-1 is a transcriptional repressor of HIV-1 LTR. Science *241*, 1481-1485.

Ahmad N, V.S. (1988). Nef protein of HIV-1 is a transcriptional repressor of HIV-1 LTR. Science *241*, 1481-1485.

Aiken C, T.D. (1995). Nef stimulates human immunodeficiency virus type 1 proviral DNA synthesis. J Virol *69*, 5048-5056.

Al-Temaimi, R.A., Jacob, S., Al-Ali, W., Thomas, D.A., and Al-Mulla, F. (2013). Reduced FHIT expression is associated with mismatch repair deficient and high CpG island methylator phenotype colorectal cancer. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society *61*, 627-638.

Alexander L, W.E., Greenough TC, Gaddis NC, Auerbach MR, Malim MH, O'Brien SJ, Walker BD, Sullivan JL, Desrosiers RC. (2000). Unusual polymorphisms in human immunodeficiency virus type 1 associated with nonprogressive infection. J Virol *74*, 4361-4376.

Allan, J.S., Coligan, J.E., Barin, F., McLane, M.F., Sodroski, J.G., Rosen, C.A., Haseltine, W.A., Lee, T.H., and Essex, M. (1985a). Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. Science *228*, 1091-1094.

Allan, J.S., Coligan, J.E., Lee, T.H., McLane, M.F., Kanki, P.J., Groopman, J.E., and Essex, M. (1985b). A new HTLV-III/LAV encoded antigen detected by antibodies from AIDS patients. Science *230*, 810-813.

Amini, S., Khalili, K., and Sawaya, B.E. (2004a). Effect of HIV-1 Vpr on cell cycle regulators. DNA and cell biology *23*, 249-260.

Amini, S., Saunders, M., Kelley, K., Khalili, K., and Sawaya, B.E. (2004b). Interplay between HIV-1 Vpr and Sp1 modulates p21(WAF1) gene expression in human astrocytes. The Journal of biological chemistry *279*, 46046-46056.

An, P., Goedert, J.J., Donfield, S., Buchbinder, S., Kirk, G.D., Detels, R., and Winkler, C.A. (2014). Regulatory Variation in HIV-1 Dependency Factor ZNRD1 Associates with Host Resistance to HIV-1 Acquisition. The Journal of infectious diseases.

An, P., and Winkler, C.A. (2010). Host genes associated with HIV/AIDS: advances in gene discovery. Trends in genetics : TIG *26*, 119-131.

Anderson, J.S., Javien, J., Nolta, J.A., and Bauer, G. (2009). Preintegration HIV-1 inhibition by a combination lentiviral vector containing a chimeric TRIM5 alpha protein, a CCR5 shRNA, and a TAR decoy. Molecular therapy : the journal of the American Society of Gene Therapy *17*, 2103-2114.

Antoni BA, R.A., Kinter A, Bodkin M, Poli G. (1994). NF-kappa B-dependent and -independent pathways of HIV activation in a chronically infected T cell line. Virology *202*, 684-694.

Appay V, F.S., Katlama C, Ait-Mohand H, Schneider L, Guihot A, Keller M, Grubeck-Loebenstein B, Simon A, Lambotte O, Hunt PW, Deeks SG, Costagliola D, Autran B, Sauce D. (2011). Old age and anti-cytomegalovirus immunity are associated with altered T-cell reconstitution in HIV-1-infected patients. Aids *25*, 1813-1822.

Archibald, D.W., Zon, L., Groopman, J.E., McLane, M.F., and Essex, M. (1986). Antibodies to human T-lymphotropic virus type III (HTLV-III) in saliva of acquired immunodeficiency syndrome (AIDS) patients and in persons at risk for AIDS. Blood *67*, 831-834.

Archin NM, K.K., Espeseth A, Dang H, Hazuda DJ, Margolis DM (2009). Expression of latent human immunodeficiency type 1 is induced by novel and selective histone deacetylase inhibitors. Aids *23*, 1799-1806.

Bailey, J.R., Sedaghat, A.R., Kieffer, T., Brennan, T., Lee, P.K., Wind-Rotolo, M., Haggerty, C.M., Kamireddi, A.R., Liu, Y., Lee, J., *et al*. (2006). Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells. Journal of virology *80*, 6441-6457.

Bangsberg, D.R., Hecht, F.M., Charlebois, E.D., Zolopa, A.R., Holodniy, M., Sheiner, L., Bamberger, J.D., Chesney, M.A., and Moss, A. (2000). Adherence to protease inhibitors, HIV-1 viral load, and development of drug resistance in an indigent population. Aids *14*, 357-366.

Barin, F., McLane, M.F., Allan, J.S., Lee, T.H., Groopman, J.E., and Essex, M. (1985). Virus envelope protein of HTLV-III represents major target antigen for antibodies in AIDS patients. Science *228*, 1094-1096.

Barnitz, R.A., Wan, F., Tripuraneni, V., Bolton, D.L., and Lenardo, M.J. (2010). Protein kinase A phosphorylation activates Vpr-induced cell cycle arrest during human immunodeficiency virus type 1 infection. Journal of virology *84*, 6410-6424.

Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., *et al*. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science *220*, 868-871.

Bartha, I., Carlson, J.M., Brumme, C.J., McLaren, P.J., Brumme, Z.L., John, M., Haas, D.W., Martinez-Picado, J., Dalmau, J., Lopez-Galindez, C., *et al*. (2013). A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. eLife *2*, e01123.

Baum, M.K., Campa, A., Lai, S., Sales Martinez, S., Tsalaile, L., Burns, P., Farahani, M., Li, Y., van Widenfelt, E., Page, J.B., *et al*. (2013). Effect of micronutrient supplementation on disease progression in asymptomatic, antiretroviral-naive, HIV-infected adults in Botswana: a randomized clinical trial. JAMA : the journal of the American Medical Association *310*, 2154-2163.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological), 289-300.

Bergonzini, V., Calistri, A., Salata, C., Del Vecchio, C., Sartori, E., Parolin, C., and Palu, G. (2009). Nef and cell signaling transduction: a possible involvement in the pathogenesis of human immunodeficiency virus-associated dementia. Journal of neurovirology *15*, 238-248.

Blagoveshchenskaya, A.D., Thomas, L., Feliciangeli, S.F., Hung, C.H., and Thomas, G. (2002). HIV-1 Nef downregulates MHC-I by a PACS-1- and PI3K-regulated ARF6 endocytic pathway. Cell *111*, 853-866.

Blankson JN, B.J., Thayil S, Yang HC, Lassen K, Lai J, Gandhi SK, Siliciano JD, Williams TM, Siliciano RF. (2007). Isolation and characterization of replication-competent human immunodeficiency virus type 1 from a subset of elite suppressors. J Virol *81*, 2508-2518.

Brass, A.L., Dykxhoorn, D.M., Benita, Y., Yan, N., Engelman, A., Xavier, R.J., Lieberman, J., and Elledge, S.J. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. Science *319*, 921-926.

Brooks, D.G., Hamer, D.H., Arlen, P.A., Gao, L., Bristol, G., Kitchen, C.M., Berger, E.A., and Zack, J.A. (2003). Molecular characterization, reactivation, and depletion of latent HIV. Immunity *19*, 413-423.

Bush, W.S., and Moore, J.H. (2012). Chapter 11: Genome-wide association studies. PLoS computational biology *8*, e1002822.

Cain, K., Bratton, S.B., Langlais, C., Walker, G., Brown, D.G., Sun, X.M., and Cohen, G.M. (2000). Apaf-1 oligomerizes into biologically active approximately 700-kDa and inactive approximately 1.4-MDa apoptosome complexes. The Journal of biological chemistry *275*, 6067-6070.

Callen BP, S.K., Egan JB. (2004). Transcriptional interference between convergent promoters caused by elongation over the promoter. Mol Cell *14*, 647-656.

Carter CC, O.-N.A., McNamara LA, Riddell J 4th, Bixby D, Savona MR, Collins KL (2010). HIV-1 infects multipotent progenitor cells causing cell death and establishing latent cellular reservoirs. Nature medicine *16*, 5.

Caserta, T.M., Smith, A.N., Gultice, A.D., Reedy, M.A., and Brown, T.L. (2003). Q-VD-OPh, a broad spectrum caspase inhibitor with potent antiapoptotic properties. Apoptosis : an international journal on programmed cell death *8*, 345-352.

Cecener, G., Tunca, B., Egeli, U., Bekar, A., Guler, G., Tolunay, S., and Aksoy, K. (2010). FHIT gene sequence variants and reduced Fhit protein expression in glioblastoma multiforme. Cellular and molecular neurobiology *30*, 301-307.

Cerdan, C., Hong, S.H., and Bhatia, M. (2007). Formation and hematopoietic differentiation of human embryoid bodies by suspension and hanging drop cultures. Curr Protoc Stem Cell Biol *Chapter 1*, Unit 1D 2.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell *147*, 358-369.

Chen, T., and Dent, S.Y. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. Nature reviews Genetics *15*, 93-106.

Cherrier, T., Suzanne, S., Redel, L., Calao, M., Marban, C., Samah, B., Mukerjee, R., Schwartz, C., Gras, G., Sawaya, B.E.*, et al*. (2009). p21(WAF1) gene promoter is epigenetically silenced by CTIP2 and SUV39H1. Oncogene *28*, 3380-3389.

Chinn, L.W., Tang, M., Kessing, B.D., Lautenberger, J.A., Troyer, J.L., Malasky, M.J., McIntosh, C., Kirk, G.D., Wolinsky, S.M., Buchbinder, S.P.*, et al*. (2010). Genetic associations of variants in genes encoding HIV-dependency factors required for HIV-1 infection. The Journal of infectious diseases *202*, 1836-1845.

Chomont, N., El-Far, M., Ancuta, P., Trautmann, L., Procopio, F.A., Yassine-Diab, B., Boucher, G., Boulassel, M.R., Ghattas, G., Brenchley, J.M.*, et al*. (2009). HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nature medicine *15*, 893-900.

Chowdhury, I.H., Wang, X.F., Landau, N.R., Robb, M.L., Polonis, V.R., Birx, D.L., and Kim, J.H. (2003). HIV-1 Vpr activates cell cycle inhibitor p21/Waf1/Cip1: a potential mechanism of G2/M cell cycle arrest. Virology *305*, 371-377.

Chowers MY, P.M., Spina CA, Richman DD, Guatelli JC. (1995). The growth advantage conferred by HIV-1 *nef* is determined at the level of viral DNA formation and is independent of CD4 downregulation. Virology *212*, 451-457.

Chowers MY, S.C., Kwoh TJ, Fitch NJ, Richman DD, Guatelli JC. (1994). Optimal infectivity *in vitro* of human immunodeficiency virus type 1 requires an intact *nef* gene. J Virol *68*, 2906-2914.

Chubb, D., Weinhold, N., Broderick, P., Chen, B., Johnson, D.C., Forsti, A., Vijayakrishnan, J., Migliorini, G., Dobbins, S.E., Holroyd, A.*, et al*. (2013). Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. Nature genetics *45*, 1221-1225.

Chun, T.W., Carruth, L., Finzi, D., Shen, X., DiGiuseppe, J.A., Taylor, H., Hermankova, M., Chadwick, K., Margolick, J., Quinn, T.C.*, et al*. (1997). Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature *387*, 183-188.

Chun TW, F.D., Margolick J, Chadwick K, Schwartz D, Siliciano RF. (1995). In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. Nat Med *1*, 1284-1290.

Chun, T.W., Finzi, D., Margolick, J., Chadwick, K., Schwartz, D., and Siliciano, R.F. (1995). In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. Nature medicine *1*, 1284-1290.

Ciuffi, A., and Barr, S.D. (2011). Identification of HIV integration sites in infected host genomic DNA. Methods *53*, 39-46.

Ciuffi A, B.G., Muñoz M, Martinez R, Loeuillet C, Rehr M, Fischer M, Günthard HF, Oxenius A, Meylan P, Bonhoeffer S, Trono D, Telenti A. (2004). Entry and transcription as key determinants of differences in CD4 T-cell permissiveness to human immunodeficiency virus type 1 infection. J Virol *78*, 10747-10754.

Cohen GB, G.R., Davis DM, Mandelboim O, Chen BK, Strominger JL, Baltimore D. (1999). The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. Immunity *10*, 661-671.

Cohen, M.S., Chen, Y.Q., McCauley, M., Gamble, T., Hosseinipour, M.C., Kumarasamy, N., Hakim, J.G., Kumwenda, J., Grinsztejn, B., Pilotto, J.H.*, et al*. (2011a). Prevention of HIV-1 infection with early antiretroviral therapy. The New England journal of medicine *365*, 493-505.

Cohen, M.S., Shaw, G.M., McMichael, A.J., and Haynes, B.F. (2011b). Acute HIV-1 Infection. The New England journal of medicine *364*, 1943-1954.

Colombo, S., Rauch, A., Rotger, M., Fellay, J., Martinez, R., Fux, C., Thurnheer, C., Gunthard, H.F., Goldstein, D.B., Furrer, H.*, et al*. (2008). The HCP5 single-nucleotide

polymorphism: a simple screening tool for prediction of hypersensitivity reaction to abacavir. The Journal of infectious diseases *198*, 864-867.

Corthésy B, K.P. (1994). Purification by DNA affinity chromatography of two polypeptides that contact the NF-AT DNA binding site in the interleukin 2 promoter. J Biol Chem *269*, 20682-20690.

Crampton N, B.W., Kirkham J, Rivetti C, Thomson NH. (2006). Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. Nucleic Acids Res *34*, 5416-5425.

Cui, K., Zang, C., Roh, T.Y., Schones, D.E., Childs, R.W., Peng, W., and Zhao, K. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell stem cell *4*, 80-93.

Daar, E.S., Moudgil, T., Meyer, R.D., and Ho, D.D. (1991). Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. The New England journal of medicine *324*, 961-964.

Dahiya, S., Liu, Y., Nonnemacher, M.R., Dampier, W., and Wigdahl, B. (2014). CCAAT enhancer binding protein and nuclear factor of activated T cells regulate HIV-1 LTR via a novel conserved downstream site in cells of the monocyte-macrophage lineage. PloS one *9*, e88116.

Dalmasso, C., Carpentier, W., Meyer, L., Rouzioux, C., Goujard, C., Chaix, M.L., Lambotte, O., Avettand-Fenoel, V., Le Clerc, S., de Senneville, L.D., *et al*. (2008). Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. PloS one *3*, e3907.

Dalmasso C, C.W., Meyer L, Rouzioux C, Goujard C, Chaix ML, Lambotte O, Avettand-Fenoel V, Le Clerc S, de Senneville LD, Deveau C, Boufassa F, Debré P, Delfraissy JF, Broet P, Theodorou I; ANRS Genome Wide Association 01. (2008). Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. PloS one *3*, e3907.

Davey, R.T., Jr., Bhat, N., Yoder, C., Chun, T.W., Metcalf, J.A., Dewar, R., Natarajan, V., Lempicki, R.A., Adelsberger, J.W., Miller, K.D., *et al*. (1999). HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. Proceedings of the National Academy of Sciences of the United States of America *96*, 15109-15114.

Degterev, A., Huang, Z., Boyce, M., Li, Y., Jagtap, P., Mizushima, N., Cuny, G.D., Mitchison, T.J., Moskowitz, M.A., and Yuan, J. (2005). Chemical inhibitor of nonapoptotic cell death with therapeutic potential for ischemic brain injury. Nature chemical biology *1*, 112-119.

Deshmane, S.L., Mukerjee, R., Fan, S., Del Valle, L., Michiels, C., Sweet, T., Rom, I., Khalili, K., Rappaport, J., Amini, S., *et al*. (2009). Activation of the oxidative stress

pathway by HIV-1 Vpr leads to induction of hypoxia-inducible factor 1alpha expression. The Journal of biological chemistry *284*, 11364-11373.

Donaghy, H., Pozniak, A., Gazzard, B., Qazi, N., Gilmour, J., Gotch, F., and Patterson, S. (2001). Loss of blood CD11c(+) myeloid and CD11c(-) plasmacytoid dendritic cells in patients with HIV-1 infection correlates with HIV-1 RNA virus load. Blood *98*, 2574-2576.

Doulatov, S., Vo, L.T., Chou, S.S., Kim, P.G., Arora, N., Li, H., Hadland, B.K., Bernstein, I.D., Collins, J.J., Zon, L.I., *et al*. (2013). Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. Cell stem cell *13*, 459-470.

du Chene, I., Basyuk, E., Lin, Y.L., Triboulet, R., Knezevich, A., Chable-Bessia, C., Mettling, C., Baillat, V., Reynes, J., Corbeau, P., *et al*. (2007). Suv39H1 and HP1gamma are responsible for chromatin-mediated HIV-1 transcriptional silencing and post-integration latency. The EMBO journal *26*, 424-435.

du Chéné I, B.E., Lin YL, Triboulet R, Knezevich A, Chable-Bessia C, Mettling C, Baillat V, Reynes J, Corbeau P, Bertrand E, Marcello A, Emiliani S, Kiernan R, Benkirane M. (2007). Suv39H1 and HP1gamma are responsible for chromatin-mediated HIV-1 transcriptional silencing and post-integration latency. EMBO J *26*, 11.

Duan, L., Ozaki, I., Oakes, J.W., Taylor, J.P., Khalili, K., and Pomerantz, R.J. (1994). The tumor suppressor protein p53 strongly alters human immunodeficiency virus type 1 replication. Journal of virology *68*, 4302-4313.

Dybul, M., Fauci, A.S., Bartlett, J.G., Kaplan, J.E., Pau, A.K., and Panel on Clinical Practices for the Treatment of, H.I.V. (2002). Guidelines for using antiretroviral agents among HIV-infected adults and adolescents. Recommendations of the Panel on Clinical Practices for Treatment of HIV. MMWR Recommendations and reports : Morbidity and mortality weekly report Recommendations and reports / Centers for Disease Control *51*, 1-55.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., *et al*. (2008). Global transcription in pluripotent embryonic stem cells. Cell stem cell *2*, 437-447.

Ellery, P.J., Tippett, E., Chiu, Y.L., Paukovics, G., Cameron, P.U., Solomon, A., Lewin, S.R., Gorry, P.R., Jaworowski, A., Greene, W.C., *et al*. (2007). The CD16+ monocyte subset is more permissive to infection and preferentially harbors HIV-1 *in vivo*. Journal of immunology *178*, 6581-6589.

Emert-Sedlak L, K.T., Lerner EC, Dai W, Foster C, Day BW, Lazo JS, Smithgall TE. (2009). Chemical library screens targeting an HIV-1 accessory factor/host cell kinase complex identify novel antiretroviral compounds. ACS Chem Biol *4*, 939-947.

Farnet CM, B.F. (1997). HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes *in vitro*. Cell *88*, 483-492.

Fauci, A.S., Pantaleo, G., Stanley, S., and Weissman, D. (1996). Immunopathogenic mechanisms of HIV infection. Annals of internal medicine *124*, 654-663.

Fellay J, G.D., Shianna KV, Colombo S, Ledergerber B, Cirulli ET, Urban TJ, Zhang K, Gumbs CE, Smith JP, Castagna A, Cozzi-Lepri A, De Luca A, Easterbrook P, Günthard HF, Mallal S, Mussini C, Dalmau J, Martinez-Picado J, Miro JM, Obel N, Wolinsky SM, Martinson JJ, Detels R, Margolick JB, Jacobson LP, Descombes P, Antonarakis SE, Beckmann JS, O'Brien SJ, Letvin NL, McMichael AJ, Haynes BF, Carrington M, Feng S, Telenti A, Goldstein DB; NIAID Center for HIV/AIDS Vaccine Immunology (CHAVI). (2009). Common genetic variation and the control of HIV-1 in humans. PLoS Genet *5*, e1000791.

Fellay, J., Ge, D., Shianna, K.V., Colombo, S., Ledergerber, B., Cirulli, E.T., Urban, T.J., Zhang, K., Gumbs, C.E., Smith, J.P.*, et al*. (2009). Common genetic variation and the control of HIV-1 in humans. PLoS genetics *5*, e1000791.

Fellay, J., Shianna, K.V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A.*, et al*. (2007). A whole-genome association study of major determinants for host control of HIV-1. Science *317*, 944-947.

Fellay J, S.K., Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL, McMichael AJ, Haynes BF, Telenti A, Goldstein DB. (2007). A whole-genome association study of major determinants for host control of HIV-1. Science *317*, 944-947.

Finzi, D., Blankson, J., Siliciano, J.D., Margolick, J.B., Chadwick, K., Pierson, T., Smith, K., Lisziewicz, J., Lori, F., Flexner, C.*, et al*. (1999). Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nature medicine *5*, 512-517.

Folks TM, C.K., Justement J, Rabson A, Duh E, Kehrl JH, Fauci AS. (1989). Tumor necrosis factor alpha induces expression of human immunodeficiency virus in a chronically infected T-cell clone. Proc Natl Acad Sci U S A *86*, 2365-2368.

Folks TM, J.J., Kinter A, Dinarello CA, Fauci AS. (1987). Cytokine-induced expression of HIV-1 in a chronically infected promonocyte cell line. Science *238*, 800-802.

Foster, J.L., Denial, S.J., Temple, B.R., and Garcia, J.V. (2011). Mechanisms of HIV-1 Nef function and intracellular signaling. Journal of neuroimmune pharmacology : the official journal of the Society on NeuroImmune Pharmacology *6*, 230-246.

Francis, A.C., Di Primio, C., Allouch, A., and Cereseto, A. (2011). Role of phosphorylation in the nuclear biology of HIV-1. Current medicinal chemistry *18*, 2904-2912.

Friedman J, C.W., Chu CK, Keedy KS, Archin NM, Margolis DM, Karn J. (2011). Epigenetic silencing of HIV-1 by the histone H3 lysine 27 methyltransferase enhancer of Zeste 2. J Virol *85*, 9078-9089.

Galic, Z., Kitchen, S.G., Kacena, A., Subramanian, A., Burke, B., Cortado, R., and Zack, J.A. (2006). T lineage differentiation from human embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *103*, 11742-11747.

Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J., Safai, B.*, et al*. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science *224*, 500-503.

Garber ME, W.P., Jones KA. (1998). HIV-1 Tat interacts with cyclin T1 to direct the P-TEFb CTD kinase complex to TAR RNA. Cold Spring Harb Symp Quant Biol *63*, 371-380.

Garcia-Perez JL, M.M., Scheys JO, Kulpa DA, Morell S, Carter CC, Hammer GD, Collins KL, O'Shea KS, Menendez P, Moran JV. (2010). Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. Nature *466*, 769-773.

Gareth M. James, H.T., Sugar CA (2000). Principal component models for sparse functional data. Biometrika, 587-602.

Garrido, N., Remohi, J., Pellicer, A., and Meseguer, M. (2006). The effectiveness of modified sperm washes in severely oligoasthenozoospermic men infected with human immunodeficiency and hepatitis C viruses. Fertility and sterility *86*, 1544-1546.

Gee, K., Angel, J.B., Mishra, S., Blahoianu, M.A., and Kumar, A. (2007). IL-10 regulation by HIV-Tat in primary human monocytic cells: involvement of calmodulin/calmodulin-dependent protein kinase-activated p38 MAPK and Sp-1 and CREB-1 transcription factors. Journal of immunology *178*, 798-807.

Geretti, A.M. (2006). HIV-1 subtypes: epidemiology and significance for HIV management. Current opinion in infectious diseases *19*, 1-7.

Geyer, M., Fackler, O.T., and Peterlin, B.M. (2001). Structure--function relationships in HIV-1 Nef. EMBO reports *2*, 580-585.

Gonzalez, E., Punzon, C., Gonzalez, M., and Fresno, M. (2001). HIV-1 Tat inhibits IL-2 gene transcription through qualitative and quantitative alterations of the cooperative Rel/AP1 complex bound to the CD28RE/AP1 composite element of the IL-2 promoter. Journal of immunology *166*, 4560-4569.

Gordon S, A.G., Garban H, Bonavida B. (2006). Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. Oncogene *25*, 17.

Government-of-Botswana (2012). 2012 Botswana National HIV&AIDS Treatment Guidelines (Gaborone Botswana).

Grewal SI, M.D. (2003). Heterochromatin and epigenetic control of gene expression. Science *301*, 798-802.

Guan, S.P., Mok, Y.K., Koo, K.N., Chu, K.L., and Wong, W.S. (2009). Chitinases: biomarkers for human diseases. Protein and peptide letters *16*, 490-498.

Gulick, R.M., Mellors, J.W., Havlir, D., Eron, J.J., Gonzalez, C., McMahon, D., Richman, D.D., Valentine, F.T., Jonas, L., Meibohm, A., *et al*. (1997). Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. The New England journal of medicine *337*, 734-739.

Hall, P., Müller, H. G., & Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. The annals of statistics, 1493-1517.

Hammer, S.M., Squires, K.E., Hughes, M.D., Grimes, J.M., Demeter, L.M., Currier, J.S., Eron, J.J., Jr., Feinberg, J.E., Balfour, H.H., Jr., Deyton, L.R., *et al*. (1997). A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. The New England journal of medicine *337*, 725-733.

Han Y, L.K., Monie D, Sedaghat AR, Shimoji S, Liu X, Pierson TC, Margolick JB, Siliciano RF, Siliciano JD. (2004). Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. J Virol *78*, 6122-6133.

Han Y, L.Y., An W, Xu J, Yang HC, O'Connell K, Dordai D, Boeke JD, Siliciano JD, Siliciano RF. (2008). Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. Cell Host Microbe *4*, 134-146.

Han Y, S.R. (2007a). Keeping quiet: microRNAs in HIV-1 latency. Nature medicine *13*, 1138-1140.

Han Y, W.-R.M., Yang HC, Siliciano JD, Siliciano RF. (2007b). Experimental approaches to the study of HIV-1 latency. Nat Rev Microbiol *5*, 95-106.

Hanna, J.H., Saha, K., and Jaenisch, R. (2010). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. Cell *143*, 508-525.

Hara, H., Friedlander, R.M., Gagliardini, V., Ayata, C., Fink, K., Huang, Z., Shimizu-Sasamata, M., Yuan, J., and Moskowitz, M.A. (1997). Inhibition of interleukin 1beta converting enzyme family proteases reduces ischemic and excitotoxic neuronal damage.

Proceedings of the National Academy of Sciences of the United States of America *94*, 2007-2012.

Harrison, K.M., Song, R., and Zhang, X. (2010). Life expectancy after HIV diagnosis based on national HIV surveillance data from 25 states, United States. Journal of acquired immune deficiency syndromes *53*, 124-130.

Hasegawa, A., Liu, H., Ling, B., Borda, J.T., Alvarez, X., Sugimoto, C., Vinet-Oliphant, H., Kim, W.K., Williams, K.C., Ribeiro, R.M., *et al*. (2009). The level of monocyte turnover predicts disease progression in the macaque model of AIDS. Blood *114*, 2917-2925.

Haseltine, W.A., and Patarca, R. (1986). AIDS virus and scrapie agent share protein. Nature *323*, 115-116.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., *et al*. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell stem cell *6*, 479-491.

Henrich, T.J., Hu, Z., Li, J.Z., Sciaranghella, G., Busch, M.P., Keating, S.M., Gallien, S., Lin, N.H., Giguel, F.F., Lavoie, L., *et al*. (2013). Long-term reduction in peripheral blood HIV type 1 reservoirs following reduced-intensity conditioning allogeneic stem cell transplantation. The Journal of infectious diseases *207*, 1694-1702.

Herbeck, J.T., Gottlieb, G.S., Winkler, C.A., Nelson, G.W., An, P., Maust, B.S., Wong, K.G., Troyer, J.L., Goedert, J.J., Kessing, B.D., *et al*. (2010). Multistage genomewide association study identifies a locus at 1q41 associated with rate of HIV-1 disease progression to clinical AIDS. The Journal of infectious diseases *201*, 618-626.

Hermankova, M., Siliciano, J.D., Zhou, Y., Monie, D., Chadwick, K., Margolick, J.B., Quinn, T.C., and Siliciano, R.F. (2003). Analysis of human immunodeficiency virus type 1 gene expression in latently infected resting CD4+ T lymphocytes *in vivo*. Journal of virology *77*, 7383-7392.

Hirsch, M.S., Brun-Vezinet, F., D'Aquila, R.T., Hammer, S.M., Johnson, V.A., Kuritzkes, D.R., Loveday, C., Mellors, J.W., Clotet, B., Conway, B., *et al*. (2000). Antiretroviral drug resistance testing in adult HIV-1 infection: recommendations of an International AIDS Society-USA Panel. JAMA : the journal of the American Medical Association *283*, 2417-2426.

Hiyoshi, M., Takahashi-Makise, N., Yoshidomi, Y., Chutiwitoonchai, N., Chihara, T., Okada, M., Nakamura, N., Okada, S., and Suzu, S. (2012). HIV-1 Nef perturbs the function, structure, and signaling of the Golgi through the Src kinase Hck. Journal of cellular physiology *227*, 1090-1097.

Huang J, W.F., Argyris E, Chen K, Liang Z, Tian H, Huang W, Squires K, Verlinghieri G, Zhang H. (2007). Cellular microRNAs contribute to HIV-1 latency in resting primary CD4+ T lymphocytes. Nat Med *13*, 1241-1247.

Huang, Y., Paxton, W.A., Wolinsky, S.M., Neumann, A.U., Zhang, L., He, T., Kang, S., Ceradini, D., Jin, Z., Yazdanbakhsh, K., *et al*. (1996). The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nature medicine *2*, 1240-1243.

Huthoff H, T.G. (2008). Restriction of retroviral replication by APOBEC3G/F and TRIM5α. Trends Microbiol *16*, 612-619.

Igarashi, T., Brown, C.R., Endo, Y., Buckler-White, A., Plishka, R., Bischofberger, N., Hirsch, V., and Martin, M.A. (2001). Macrophage are the principal reservoir and sustain high virus loads in rhesus macaques after the depletion of CD4+ T cells by a highly pathogenic simian immunodeficiency virus/HIV type 1 chimera (SHIV): Implications for HIV-1 infections of humans. Proceedings of the National Academy of Sciences of the United States of America *98*, 658-663.

Imamichi, H., Crandall, K.A., Natarajan, V., Jiang, M.K., Dewar, R.L., Berg, S., Gaddam, A., Bosche, M., Metcalf, J.A., Davey, R.T., Jr., *et al*. (2001). Human immunodeficiency virus type 1 quasi species that rebound after discontinuation of highly active antiretroviral therapy are similar to the viral quasi species present before initiation of therapy. The Journal of infectious diseases *183*, 36-50.

International, H.I.V.C.S., Pereyra, F., Jia, X., McLaren, P.J., Telenti, A., de Bakker, P.I., Walker, B.D., Ripke, S., Brumme, C.J., Pulit, S.L., *et al*. (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science *330*, 1551-1557.

Jacque JM, S.M. (2006). The inner-nuclear-envelope protein emerin regulates HIV-1 infectivity. Nature *441*, 641-645.

Jager, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K., *et al*. (2012). Global landscape of HIV-human protein complexes. Nature *481*, 365-370.

James, G.M. (2002). Generalized linear models with functional predictor variables. Journal of the Royal Statistical Society Series B *64*, 411–432.

Jaworowski, A., Kamwendo, D.D., Ellery, P., Sonza, S., Mwapasa, V., Tadesse, E., Molyneux, M.E., Rogerson, S.J., Meshnick, S.R., and Crowe, S.M. (2007). CD16+ monocyte subset preferentially harbors HIV-1 and is expanded in pregnant Malawian women with Plasmodium falciparum malaria and HIV-1 infection. The Journal of infectious diseases *196*, 38-42.

Jeon, B.N., Kim, M.K., Choi, W.I., Koh, D.I., Hong, S.Y., Kim, K.S., Kim, M., Yun, C.O., Yoon, J., Choi, K.Y., *et al*. (2012a). KR-POK interacts with p53 and represses its ability to activate transcription of p21WAF1/CDKN1A. Cancer research *72*, 1137-1148.

Jeon, B.N., Kim, Y.S., Choi, W.I., Koh, D.I., Kim, M.K., Yoon, J.H., Kim, M.Y., Hur, B., Paik, P.D., and Hur, M.W. (2012b). Kr-pok increases FASN expression by modulating

the DNA binding of SREBP-1c and Sp1 at the proximal promoter. Journal of lipid research *53*, 755-766.

Jones KA, K.J., Luciw PA, Tjian R. (1986). Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. Science *233*, 755-759.

Jordan A, D.P., Verdin E (2001). The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. EMBO J *20*, 1726-1738.

Kagi, D., Vignaux, F., Ledermann, B., Burki, K., Depraetere, V., Nagata, S., Hengartner, H., and Golstein, P. (1994). Fas and perforin pathways as major mechanisms of T cell-mediated cytotoxicity. Science *265*, 528-530.

Kaiser, W.J., Upton, J.W., Long, A.B., Livingston-Rosanoff, D., Daley-Bauer, L.P., Hakem, R., Caspary, T., and Mocarski, E.S. (2011). RIP3 mediates the embryonic lethality of caspase-8-deficient mice. Nature *471*, 368-372.

Kamata, M., Liu, S., Liang, M., Nagaoka, Y., and Chen, I.S. (2010). Generation of human induced pluripotent stem cells bearing an anti-HIV transgene by a lentiviral vector carrying an internal murine leukemia virus promoter. Human gene therapy *21*, 1555-1567.

Kamine, J., Subramanian, T., and Chinnadurai, G. (1991). Sp1-dependent activation of a synthetic promoter by human immunodeficiency virus type 1 Tat protein. Proceedings of the National Academy of Sciences of the United States of America *88*, 8510-8514.

Kao, S.Y., Calman, A.F., Luciw, P.A., and Peterlin, B.M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by *tat* gene product. Nature *330*, 489-493.

Kapasi, A.A., Fan, S., and Singhal, P.C. (2006). Scatter factor mitigates HIV-1 gp120-induced human mesangial cell injury. Nephron Experimental nephrology *103*, e103-108.

Karolchik, D., Hinrichs, A.S., and Kent, W.J. (2009). The UCSC Genome Browser. Curr Protoc Bioinformatics *Chapter 1*, Unit1 4.

Kestler HW 3rd, R.D., Mori K, Panicali DL, Sehgal PK, Daniel MD, Desrosiers RC. (1991). Importance of the *nef* gene for maintenance of high virus loads and for development of AIDS. Cell *65*, 651-662.

Kestler, H.W., 3rd, Ringler, D.J., Mori, K., Panicali, D.L., Sehgal, P.K., Daniel, M.D., and Desrosiers, R.C. (1991). Importance of the *nef* gene for maintenance of high virus loads and for development of AIDS. Cell *65*, 651-662.

Kim, L.U., Johnson, M.R., Barton, S., Nelson, M.R., Sontag, G., Smith, J.R., Gotch, F.M., and Gilmour, J.W. (1999). Evaluation of sperm washing as a potential method of reducing HIV transmission in HIV-discordant couples wishing to have children. Aids *13*, 645-651.

Kim YK, B.C., Pearson R, Tyagi M, West MJ, Wong J, Wu SY, Chiang CM, Karn J. (2006). Recruitment of TFIIH to the HIV LTR is a rate-limiting step in the emergence of HIV from latency. EMBO J *25*, 3596-3604.

Kino, T., Tsukamoto, M., and Chrousos, G. (2002). Transcription factor TFIIH components enhance the GR coactivator activity but not the cell cycle-arresting activity of the human immunodeficiency virus type-1 protein Vpr. Biochemical and biophysical research communications *298*, 17-23.

Kitchen, L.W., Barin, F., Sullivan, J.L., McLane, M.F., Brettler, D.B., Levine, P.H., and Essex, M. (1984). Aetiology of AIDS--antibodies to human T-cell leukaemia virus (type III) in haemophiliacs. Nature *312*, 367-369.

Kuehn, B.M. (2006). UNAIDS report: AIDS epidemic slowing, but huge challenges remain. JAMA : the journal of the American Medical Association *296*, 29-30.

Kumar, P. (2013). Long term non-progressor (LTNP) HIV infection. The Indian journal of medical research *138*, 291-293.

Lahti, A.L., Manninen, A., and Saksela, K. (2003). Regulation of T cell activation by HIV-1 accessory proteins: Vpr acts via distinct mechanisms to cooperate with Nef in NFAT-directed gene expression and to promote transactivation by CREB. Virology *310*, 190-196.

Laird, N.M., & Ware, J. H. (1982). Random-effects models for longitudinal data. Biometrics, 963-974.

Lambotte, O., Taoufik, Y., de Goer, M.G., Wallon, C., Goujard, C., and Delfraissy, J.F. (2000). Detection of infectious HIV in circulating monocytes from patients on prolonged highly active antiretroviral therapy. Journal of acquired immune deficiency syndromes *23*, 114-119.

Lassen, K.G., Bailey, J.R., and Siliciano, R.F. (2004). Analysis of human immunodeficiency virus type 1 transcriptional elongation in resting CD4+ T cells *in vivo*. Journal of virology *78*, 9105-9114.

Lavallard, V.J., Pradelli, L.A., Paul, A., Beneteau, M., Jacquel, A., Auberger, P., and Ricci, J.E. (2009). Modulation of caspase-independent cell death leads to resensitization of imatinib mesylate-resistant cells. Cancer research *69*, 3013-3020.

Le Clerc, S., Limou, S., Coulonges, C., Carpentier, W., Dina, C., Taing, L., Delaneau, O., Labib, T., Sladek, R., Group, A.G., *et al*. (2009). Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS Genomewide Association Study 03). The Journal of infectious diseases *200*, 1194-1201.

Learmont, J.C., Geczy, A.F., Mills, J., Ashton, L.J., Raynes-Greenow, C.H., Garsia, R.J., Dyer, W.B., McIntyre, L., Oelrichs, R.B., Rhodes, D.I., *et al*. (1999). Immunologic and virologic status after 14 to 18 years of infection with an attenuated strain of HIV-1. A

report from the Sydney Blood Bank Cohort. The New England journal of medicine *340*, 1715-1722.

Lee, T.H., Coligan, J.E., Allan, J.S., McLane, M.F., Groopman, J.E., and Essex, M. (1986). A new HTLV-III/LAV protein encoded by a gene found in cytopathic retroviruses. Science *231*, 1546-1549.

Lee, T.H., Coligan, J.E., and Essex, M. (1984a). Human T-cell leukemia virus specific antigens. Princess Takamatsu symposia *15*, 197-203.

Lee, T.H., Coligan, J.E., Sodroski, J.G., Haseltine, W.A., Salahuddin, S.Z., Wong-Staal, F., Gallo, R.C., and Essex, M. (1984b). Antigens encoded by the 3'-terminal region of human T-cell leukemia virus: evidence for a functional gene. Science *226*, 57-61.

Lenasi T, C.X., Peterlin BM. (2008). Transcriptional interference antagonizes proviral gene expression to promote HIV latency. Cell Host Microbe *4*, 123-133.

Lenassi, M., Cagney, G., Liao, M., Vaupotic, T., Bartholomeeusen, K., Cheng, Y., Krogan, N.J., Plemenitas, A., and Peterlin, B.M. (2010). HIV Nef is secreted in exosomes and triggers apoptosis in bystander CD4+ T cells. Traffic *11*, 110-122.

Leonard JA, F.T., Carter CC, Schaefer M, Collins KL. (2011). HIV-1 Nef disrupts intracellular trafficking of major histocompatibility complex class I, CD4, CD8, and CD28 by distinct pathways that share common elements. J Virology *85*, 6867-6881.

Lewinski MK, B.D., Shinn P, Chen H, Hoffmann C, Hannenhalli S, Verdin E, Berry CC, Ecker JR, Bushman FD. (2009). Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. J Virol *79*, 6610-6619.

Li, Y., Liu, M., and Yang, S.T. (2014). Dendritic cells derived from pluripotent stem cells: Potential of large scale production. World journal of stem cells *6*, 1-10.

Limou, S., Coulonges, C., Herbeck, J.T., van Manen, D., An, P., Le Clerc, S., Delaneau, O., Diop, G., Taing, L., Montes, M., *et al*. (2010). Multiple-cohort genetic association study reveals CXCR6 as a new chemokine receptor involved in long-term nonprogression to AIDS. The Journal of infectious diseases *202*, 908-915.

Limou, S., Le Clerc, S., Coulonges, C., Carpentier, W., Dina, C., Delaneau, O., Labib, T., Taing, L., Sladek, R., Deveau, C., *et al*. (2009). Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). The Journal of infectious diseases *199*, 419-426.

Lin, X. (1997). Variance component testing in generalised linear models with random effects. Biometrika *84*, 309-326.

Llano M, S.D., Meehan A, Wongthida P, Peretz M, Walker WH, Teo W, Poeschla EM. (2006). An essential role for LEDGF/p75 in HIV integration. Science *314*, 461-464.

Loetscher P, M.B., Baggiolini M. (2000). Chemokines and their receptors in lymphocyte traffic and HIV infection. Adv Immunol *74*, 127-180.

Ma, Y., Ramezani, A., Lewis, R., Hawley, R.G., and Thomson, J.A. (2003). High-level sustained transgene expression in human embryonic stem cells using lentiviral vectors. Stem Cells *21*, 111-117.

MacNeil, A., Sankale, J.L., Meloni, S.T., Sarr, A.D., Mboup, S., and Kanki, P. (2006). Genomic sites of human immunodeficiency virus type 2 (HIV-2) integration: similarities to HIV-1 *in vitro* and possible differences *in vivo*. Journal of virology *80*, 7316-7321.

Malim, M.H., and Emerman, M. (2008). HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell host & microbe *3*, 388-398.

Marban C, S.S., Dequiedt F, de Walque S, Redel L, Van Lint C, Aunis D, Rohr O. (2007). Recruitment of chromatin-modifying enzymes by CTIP2 promotes HIV-1 transcriptional silencing. EMBO J *26*, 11.

Masters, S.L., Gerlic, M., Metcalf, D., Preston, S., Pellegrini, M., O'Donnell, J.A., McArthur, K., Baldwin, T.M., Chevrier, S., Nowell, C.J., *et al*. (2012). NLRP1 inflammasome activation induces pyroptosis of hematopoietic progenitor cells. Immunity *37*, 1009-1023.

Mayte Coiras, M.R.L.-H., Mayte Pérez-Olmeda and José Alcamí (2009). Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. Nature Reviews *7*, 15.

McDonald D, V.M., Lucero G, Svitkina TM, Borisy GG, Emerman M, Hope TJ. (2002). Visualization of the intracellular behavior of HIV in living cells. J Cell Biol *159*, 441-452.

McIlroy, D., Autran, B., Cheynier, R., Wain-Hobson, S., Clauvel, J.P., Oksenhendler, E., Debre, P., and Hosmalin, A. (1995). Infection frequency of dendritic cells and CD4+ T lymphocytes in spleens of human immunodeficiency virus-positive patients. Journal of virology *69*, 4737-4745.

McLaren, P.J., Coulonges, C., Ripke, S., van den Berg, L., Buchbinder, S., Carrington, M., Cossarizza, A., Dalmau, J., Deeks, S.G., Delaneau, O., *et al*. (2013). Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. PLoS pathogens *9*, e1003515.

Migueles, S.A., Sabbaghian, M.S., Shupert, W.L., Bettinotti, M.P., Marincola, F.M., Martino, L., Hallahan, C.W., Selig, S.M., Schwartz, D., Sullivan, J., *et al*. (2000). HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proceedings of the National Academy of Sciences of the United States of America *97*, 2709-2714.

Migueles SA, S.M., Shupert WL, Bettinotti MP, Marincola FM, Martino L, Hallahan CW, Selig SM, Schwartz D, Sullivan J, Connors M. (2000). HLA B*5701 is highly associated

with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proc Natl Acad Sci U S A *97*, 2709-2714.

Miura, T., Brockman, M.A., Brumme, C.J., Brumme, Z.L., Carlson, J.M., Pereyra, F., Trocha, A., Addo, M.M., Block, B.L., Rothchild, A.C.*, et al*. (2008). Genetic characterization of human immunodeficiency virus type 1 in elite controllers: lack of gross genetic defects or common amino acid changes. Journal of virology *82*, 8422-8430.

Morgan, C.R., Miglionico, B.V., and Engen, J.R. (2011). Effects of HIV-1 Nef on human N-myristoyltransferase 1. Biochemistry *50*, 3394-3403.

Murry, C.E., and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell *132*, 661-680.

Nabel G, B.D. (1987). An inducible transcription factor activates expression of human immunodeficiency virus in T cells. Nature *326*, 711-713.

Nagpure, A., Choudhary, B., and Gupta, R.K. (2013). Chitinases: in agriculture and human healthcare. Critical reviews in biotechnology.

Nalls, M.A., Couper, D.J., Tanaka, T., van Rooij, F.J., Chen, M.H., Smith, A.V., Toniolo, D., Zakai, N.A., Yang, Q., Greinacher, A.*, et al*. (2011). Multiple loci are associated with white blood cell phenotypes. PLoS genetics *7*, e1002113.

Nekhai, S., Zhou, M., Fernandez, A., Lane, W.S., Lamb, N.J., Brady, J., and Kumar, A. (2002). HIV-1 Tat-associated RNA polymerase C-terminal domain kinase, CDK2, phosphorylates CDK7 and stimulates Tat-mediated transcription. The Biochemical journal *364*, 649-657.

Nelson, P.J., Sunamoto, M., Husain, M., and Gelman, I.H. (2002). HIV-1 expression induces cyclin D1 expression and pRb phosphorylation in infected podocytes: cell-cycle mechanisms contributing to the proliferative phenotype in HIV-associated nephropathy. BMC microbiology *2*, 26.

NIAID (2014). “Mississippi baby” now has detectable HIV, researchers find. (NIH News).

Nisole, S., Stoye, J.P., and Saib, A. (2005). TRIM family proteins: retroviral restriction and antiviral defence. Nature reviews Microbiology *3*, 799-808.

Norrman, K., Fischer, Y., Bonnamy, B., Wolfhagen Sand, F., Ravassard, P., and Semb, H. (2010). Quantitative comparison of constitutive promoters in human ES cells. PloS one *5*, e12413.

Novitsky, V.A., Gilbert, P.B., Shea, K., McLane, M.F., Rybak, N., Klein, I., Thior, I., Ndung'u, T., Lee, T.H., and Essex, M.E. (2006). Interactive association of proviral load and IFN-gamma-secreting T cell responses in HIV-1C infection. Virology *349*, 142-155.

O'Brien, S.J., and Hendrickson, S.L. (2013). Host genomic influences on HIV/AIDS. Genome biology *14*, 201.

O'Brien, S.J., and Nelson, G.W. (2004). Human genes that limit AIDS. Nature genetics *36*, 565-574.

O'Brien, T.R., Winkler, C., Dean, M., Nelson, J.A., Carrington, M., Michael, N.L., and White, G.C., 2nd (1997). HIV-1 infection in a man homozygous for CCR5 delta 32. Lancet *349*, 1219.

Omoto S, I.M., Tsutsumi Y, Ichikawa Y, Okuyama H, Brisibe EA, Saksena NK, Fujii YR. (2004). HIV-1 *nef* suppression by virally encoded microRNA. Retrovirology *1*.

Onder TT, K.N., Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcarci BO, Unternaehrer J, Gupta PB, Lander ES, Armstrong SA, Daley GQ. (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598-602.

Ozato, K., Shin, D.M., Chang, T.H., and Morse, H.C., 3rd (2008). TRIM family proteins and their emerging roles in innate immunity. Nature reviews Immunology *8*, 849-860.

Padian, N.S., Isbell, M.T., Russell, E.S., and Essex, M. (2012). The future of HIV prevention. Journal of acquired immune deficiency syndromes *60 Suppl 2*, S22-26.

Palacios, J.A., Perez-Pinar, T., Toro, C., Sanz-Minguela, B., Moreno, V., Valencia, E., Gomez-Hernando, C., and Rodes, B. (2012). Long-term nonprogressor and elite controller patients who control viremia have a higher percentage of methylation in their HIV-1 proviral promoters than aviremic patients receiving highly active antiretroviral therapy. Journal of virology *86*, 13081-13084.

Palmer, S., Maldarelli, F., Wiegand, A., Bernstein, B., Hanna, G.J., Brun, S.C., Kempf, D.J., Mellors, J.W., Coffin, J.M., and King, M.S. (2008). Low-level viremia persists for at least 7 years in patients on suppressive antiretroviral therapy. Proceedings of the National Academy of Sciences of the United States of America *105*, 3879-3884.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell stem cell *1*, 299-312.

Park IH, L.P., Zhao R, Huo H, Daley GQ. (2008). Generation of human-induced pluripotent stem cells. Nature Protocol *3*, 1180-1186.

Pasquier, C., Daudin, M., Righi, L., Berges, L., Thauvin, L., Berrebi, A., Massip, P., Puel, J., Bujan, L., and Izopet, J. (2000). Sperm washing and virus nucleic acid detection to reduce HIV and hepatitis C virus transmission in serodiscordant couples wishing to have children. Aids *14*, 2093-2099.

Pearson R, K.Y., Hokello J, Lassen K, Friedman J, Tyagi M, Karn J. (2008). Epigenetic silencing of HIV transcription by formation of restrictive chromatin structures at the viral

long terminal repeat drives the progressive entry of HIV into latency. J Virol *82*, 12291-12303.

Pelak, K., Goldstein, D.B., Walley, N.M., Fellay, J., Ge, D., Shianna, K.V., Gumbs, C., Gao, X., Maia, J.M., Cronin, K.D., *et al*. (2010). Host determinants of HIV-1 control in African Americans. The Journal of infectious diseases *201*, 1141-1149.

Persaud, D., Gay, H., Ziemniak, C., Chen, Y.H., Piatak, M., Jr., Chun, T.W., Strain, M., Richman, D., and Luzuriaga, K. (2013). Absence of detectable HIV-1 viremia after treatment cessation in an infant. The New England journal of medicine *369*, 1828-1835.

Persaud D, P.T., Ruff C, Finzi D, Chadwick KR, Margolick JB, Ruff A, Hutton N, Ray S, Siliciano RF. (2000). A stable latent reservoir for HIV-1 in resting CD4(+) T lymphocytes in infected children. J Clin Invest *105*, 995-1003.

Pierson, T., Hoffman, T.L., Blankson, J., Finzi, D., Chadwick, K., Margolick, J.B., Buck, C., Siliciano, J.D., Doms, R.W., and Siliciano, R.F. (2000). Characterization of chemokine receptor utilization of viruses in the latent reservoir for human immunodeficiency virus type 1. Journal of virology *74*, 7824-7833.

Piot, P., and Quinn, T.C. (2013). Response to the AIDS pandemic--a global health model. The New England journal of medicine *368*, 2210-2218.

Pirruccello, M., Sondermann, H., Pelton, J.G., Pellicena, P., Hoelz, A., Chernoff, J., Wemmer, D.E., and Kuriyan, J. (2006). A dimeric kinase assembly underlying autophosphorylation in the p21 activated kinases. Journal of molecular biology *361*, 312-326.

Pizzato M, P.E., Göttlinger HG (2008). Nef can enhance the infectivity of receptor-pseudotyped human immunodeficiency virus type 1 particles. J Virol *82*, 10811-10819.

Pomerantz, R.J. (2002a). Reservoirs of human immunodeficiency virus type 1: the main obstacles to viral eradication. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America *34*, 91-97.

Pomerantz, R.J. (2002b). Reservoirs of human immunodeficiency virus type 1: the main obstacles to viral eradication. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America *34*, 7.

Pomerantz, R.J., and Horn, D.L. (2003). Twenty years of therapy for HIV-1 infection. Nature medicine *9*, 867-873.

Popovic, M., Sarngadharan, M.G., Read, E., and Gallo, R.C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science *224*, 497-500.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics *38*, 904-909.

Ptak RG, G.B., Hartman TL, Watson KM, Osterling MC, Buckheit RW Jr, Townsend LB, Drach JC. (2010). Inhibition of human immunodeficiency virus type 1 by triciribine involves the accessory protein *nef*. Antimicrob Agents Chemother *54*, 1512-1519.

Quillent, C., Oberlin, E., Braun, J., Rousset, D., Gonzalez-Canali, G., Metais, P., Montagnier, L., Virelizier, J.L., Arenzana-Seisdedos, F., and Beretta, A. (1998). HIV-1-resistance phenotype conferred by combination of two separate inherited mutations of CCR5 gene. Lancet *351*, 14-18.

Ramratnam, B., Mittler, J.E., Zhang, L., Boden, D., Hurley, A., Fang, F., Macken, C.A., Perelson, A.S., Markowitz, M., and Ho, D.D. (2000). The decay of the latent reservoir of replication-competent HIV-1 is inversely correlated with the extent of residual viral replication during prolonged anti-retroviral therapy. Nature medicine *6*, 82-85.

Raposo, R.A., Abdel-Mohsen, M., Bilska, M., Montefiori, D.C., Nixon, D.F., and Pillai, S.K. (2013). Effects of cellular activation on anti-HIV-1 restriction factor expression profile in primary cells. Journal of virology *87*, 11924-11929.

Rathinam, V.A., Vanaja, S.K., and Fitzgerald, K.A. (2012). Regulation of inflammasome signaling. Nature immunology *13*, 333-342.

Rauch, S., Pulkkinen, K., Saksela, K., and Fackler, O.T. (2008). Human immunodeficiency virus type 1 Nef recruits the guanine exchange factor Vav1 via an unexpected interface into plasma membrane microdomains for association with p21-activated kinase 2 activity. Journal of virology *82*, 2918-2929.

Reuter, S., Bartelmann, M., Vogt, M., Geisen, C., Napierski, I., Kahn, T., Delius, H., Lichter, P., Weitz, S., Korn, B.*, et al*. (1998). APM-1, a novel human gene, identified by aberrant co-transcription with papillomavirus oncogenes in a cervical carcinoma cell line, encodes a BTB/POZ-zinc finger protein with growth inhibitory activity. The EMBO journal *17*, 215-222.

Rice, J.A., & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. Journal of the Royal Statistical Society Series B (Methodological), 233-243.

Rickabaugh TM, J.B. (2010). A challenge for the future: aging and HIV infection. Immunol Res *48*, 59-71.

Robinson, G.K. (1991). That BLUP is a good thing: The estimation of random effects. Statistical science, 15-32.

Rompani, S.B., and Cepko, C.L. (2008). Retinal progenitor cells can produce restricted subsets of horizontal cells. Proceedings of the National Academy of Sciences of the United States of America *105*, 192-197.

Rothenberg, R., Woelfel, M., Stoneburner, R., Milberg, J., Parker, R., and Truman, B. (1987). Survival with the acquired immunodeficiency syndrome. Experience with 5833 cases in New York City. The New England journal of medicine *317*, 1297-1302.

Ryu, J., Lee, H.J., Kim, K.A., Lee, J.Y., Lee, K.S., Park, J., and Choi, S.Y. (2004). Intracellular delivery of p53 fused to the basic domain of HIV-1 Tat. Molecules and cells *17*, 353-359.

Saez-Cirion, A., Bacchus, C., Hocqueloux, L., Avettand-Fenoel, V., Girault, I., Lecuroux, C., Potard, V., Versmisse, P., Melard, A., Prazuck, T., *et al*. (2013). Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI Study. PLoS pathogens *9*, e1003211.

Sarngadharan, M.G., Popovic, M., Bruch, L., Schupbach, J., and Gallo, R.C. (1984). Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS. Science *224*, 506-508.

Sawaya, B.E., Khalili, K., Mercer, W.E., Denisova, L., and Amini, S. (1998). Cooperative actions of HIV-1 Vpr and p53 modulate viral gene transcription. The Journal of biological chemistry *273*, 20052-20057.

Schlesinger, S., and Goff, S.P. (2013). Silencing of proviruses in embryonic cells: efficiency, stability and chromatin modifications. EMBO reports *14*, 73-79.

Schröder AR, S.P., Chen H, Berry C, Ecker JR, Bushman F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521-529.

Schuster, S.C., Miller, W., Ratan, A., Tomsho, L.P., Giardine, B., Kasson, L.R., Harris, R.S., Petersen, D.C., Zhao, F., Qi, J., *et al*. (2010). Complete Khoisan and Bantu genomes from southern Africa. Nature *463*, 943-947.

Schwartz O, M.V., Danos O, Heard JM. (1995). Human immunodeficiency virus type 1 Nef increases the efficiency of reverse transcription in the infected cell. J Virol *69*, 4053-4059.

SE, R.T., Ding J, Tien J, McClelland RS, Baeten JM, Overbaugh J. (2012). Efficient use of longitudinal CD4 counts and viral load measures in survival analysis. Stat Med *31*, 2086-2097.

Seaton, K.E., and Smith, C.D. (2008). N-Myristoyltransferase isozymes exhibit differential specificity for human immunodeficiency virus type 1 Gag and Nef. The Journal of general virology *89*, 288-296.

Senju, S., Haruta, M., Matsumura, K., Matsunaga, Y., Fukushima, S., Ikeda, T., Takamatsu, K., Irie, A., and Nishimura, Y. (2011). Generation of dendritic cells and macrophages from human induced pluripotent stem cells aiming at cell therapy. Gene therapy *18*, 874-883.

Shacklett, B.L. (2006). Understanding the "lucky few": the conundrum of HIV-exposed, seronegative individuals. Current HIV/AIDS reports *3*, 26-31.

Shan L, Y.H., Rabi SA, Bravo HC, Shroff NS, Irizarry RA, Zhang H, Margolick JB, Siliciano JD, Siliciano RF. (2011). Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. J Virol *85*, 5384-5393.

Shapiro, R.L., Kitch, D., Ogwu, A., Hughes, M.D., Lockman, S., Powis, K., Souda, S., Moffat, C., Moyo, S., McIntosh, K.*, et al*. (2013). HIV transmission and 24-month survival in a randomized trial of HAART to prevent MTCT during pregnancy and breastfeeding in Botswana. Aids *27*, 1911-1920.

Shiau, S., and Kuhn, L. (2014). Antiretroviral treatment in HIV-infected infants and young children: novel issues raised by the Mississippi baby. Expert review of anti-infective therapy *12*, 307-318.

Shirakawa, K., Takaori-Kondo, A., Yokoyama, M., Izumi, T., Matsui, M., Io, K., Sato, T., Sato, H., and Uchiyama, T. (2008). Phosphorylation of APOBEC3G by protein kinase A regulates its interaction with HIV-1 Vif. Nature structural & molecular biology *15*, 1184-1191.

Shun MC, D.J., Vandegraaff N, Engelman A. (2007a). Wild-type levels of human immunodeficiency virus type 1 infectivity in the absence of cellular emerin protein. J Virol *81*, 166-172.

Shun MC, R.N., Vandegraaff N, Daigle JE, Hughes S, Kellam P, Cherepanov P, Engelman A. (2007b). LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev *21*, 1767-1778.

Siddiqui RA, S.U., Altmüller J, Fritzer E, Nothnagel M, Dalibor N, Fellay J, Kaup FJ, Stahl-Hennig C, Nürnberg P, Krawczak M, Platzer M. (2009). X chromosomal variation is associated with slow progression to AIDS in HIV-1-infected women. Am J Hum Genet *85*, 228-239.

Silberner, J. (1985). AIDS blood screen approved. Science news *127*, 148.

Siliciano, J.D., Kajdas, J., Finzi, D., Quinn, T.C., Chadwick, K., Margolick, J.B., Kovacs, C., Gange, S.J., and Siliciano, R.F. (2003). Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. Nature medicine *9*, 727-728.

Simmons, A., Aluvihare, V., and McMichael, A. (2001). Nef triggers a transcriptional program in T cells imitating single-signal T cell activation and inducing HIV virulence mediators. Immunity *14*, 763-777.

Skibola, C.F., Bracci, P.M., Halperin, E., Conde, L., Craig, D.W., Agana, L., Iyadurai, K., Becker, N., Brooks-Wilson, A., Curry, J.D., *et al*. (2009). Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nature genetics *41*, 873-875.

Smith, E.N., Chen, W., Kahonen, M., Kettunen, J., Lehtimaki, T., Peltonen, L., Raitakari, O.T., Salem, R.M., Schork, N.J., Shaw, M., *et al*. (2010). Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. PLoS genetics *6*, e1001094.

Sonza, S., Mutimer, H.P., Oelrichs, R., Jardine, D., Harvey, K., Dunne, A., Purcell, D.F., Birch, C., and Crowe, S.M. (2001). Monocytes harbour replication-competent, non-latent HIV-1 in patients on highly active antiretroviral therapy. Aids *15*, 17-22.

Spina CA, K.T., Chowers MY, Guatelli JC, Richman DD. (1994). The importance of *nef* in the induction of human immunodeficiency virus type 1 replication from primary quiescent CD4 lymphocytes. J Exp Med *179*, 115-123.

Spivakov, M., and Fisher, A.G. (2007). Epigenetic signatures of stem-cell identity. Nature reviews Genetics *8*, 263-271.

Strain, M.C., Gunthard, H.F., Havlir, D.V., Ignacio, C.C., Smith, D.M., Leigh-Brown, A.J., Macaranas, T.R., Lam, R.Y., Daly, O.A., Fischer, M., *et al*. (2003). Heterogeneous clearance rates of long-lived lymphocytes infected with HIV: intrinsic stability predicts lifelong persistence. Proceedings of the National Academy of Sciences of the United States of America *100*, 4819-4824.

Stremlau M, O.C., Perron MJ, Kiessling M, Autissier P, Sodroski J. (2004). The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature *427*, 848-853.

Sune, C., and Garcia-Blanco, M.A. (1995). Sp1 transcription factor is required for *in vitro* basal and Tat-activated transcription from the human immunodeficiency virus type 1 long terminal repeat. Journal of virology *69*, 6572-6576.

Tang, M., Lautenberger, J.A., Gao, X., Sezgin, E., Hendrickson, S.L., Troyer, J.L., David, V.A., Guan, L., McIntosh, C.E., Guo, X., *et al*. (2012). The principal genetic determinants for nasopharyngeal carcinoma in China involve the HLA class I antigen recognition groove. PLoS genetics *8*, e1003103.

Teich NM, W.R., Martin GR, Lowy DR. (1977). Virus infection of murine teratocarcinoma stem cell lines. Cell *12*, 973-982.

Terwilliger E, S.J., Rosen CA, Haseltine WA (1986). Effects of mutations within the 3' orf open reading frame region of human T-cell lymphotropic virus type III (HTLV-III/LAV) on replication and cytopathogenicity. J Virol *60*, 754-760.

Thakur, B.K., Chandra, A., Dittrich, T., Welte, K., and Chandra, P. (2012). Inhibition of SIRT1 by HIV-1 viral protein Tat results in activation of p53 pathway. Biochemical and biophysical research communications *424*, 245-250.

Thornberry, N.A., Bull, H.G., Calaycay, J.R., Chapman, K.T., Howard, A.D., Kostura, M.J., Miller, D.K., Molineaux, S.M., Weidner, J.R., Aunins, J., *et al*. (1992). A novel heterodimeric cysteine protease is required for interleukin-1 beta processing in monocytes. Nature *356*, 768-774.

Troyer, J.L., Nelson, G.W., Lautenberger, J.A., Chinn, L., McIntosh, C., Johnson, R.C., Sezgin, E., Kessing, B., Malasky, M., Hendrickson, S.L., *et al*. (2011). Genome-wide association study implicates PARD3B-based AIDS restriction. The Journal of infectious diseases *203*, 1491-1502.

Tyagi M, K.J. (2007). CBF-1 promotes transcriptional silencing during the establishment of HIV-1 latency. EMBO J *26*, 10.

Tyagi M, P.R., Karn J. (2010). Establishment of HIV latency in primary CD4+ cells is due to epigenetic transcriptional silencing and P-TEFb restriction. J Virology *84*, 6425-6437.

UNAIDS (2012). UNAIDS world AIDS day 2012 report. .

UNAIDS (2013a). UNAIDS Botswana.

UNAIDS (2013b). UNAIDS report on the global AIDS epidemic 2013.

Upton, J.W., and Chan, F.K. (2014). Staying alive: cell death in antiviral immunity. Molecular cell *54*, 273-280.

van Manen, D., Delaneau, O., Kootstra, N.A., Boeser-Nunnink, B.D., Limou, S., Bol, S.M., Burger, J.A., Zwinderman, A.H., Moerland, P.D., van 't Slot, R., *et al*. (2011). Genome-wide association scan in HIV-1-infected individuals identifying variants influencing disease course. PloS one *6*, e22208.

van Manen D, K.N., Boeser-Nunnink B, Handulle MA, van't Wout AB, Schuitemaker H. (2009). Association of HLA-C and HCP5 gene regions with the clinical course of HIV-1 infection. Aids *23*, 19-28.

van Manen, D., Kootstra, N.A., Boeser-Nunnink, B., Handulle, M.A., van't Wout, A.B., and Schuitemaker, H. (2009). Association of HLA-C and HCP5 gene regions with the clinical course of HIV-1 infection. Aids *23*, 19-28.

van Manen, D., van 't Wout, A.B., and Schuitemaker, H. (2012). Genome-wide association studies on HIV susceptibility, pathogenesis and pharmacogenomics. Retrovirology *9*, 70.

van Wilgenburg, B., Browne, C., Vowles, J., and Cowley, S.A. (2013). Efficient, long term production of monocyte-derived macrophages from human pluripotent stem cells under partly-defined and fully-defined conditions. PloS one *8*, e71098.

Veazey, R.S., DeMaria, M., Chalifoux, L.V., Shvetz, D.E., Pauley, D.R., Knight, H.L., Rosenzweig, M., Johnson, R.P., Desrosiers, R.C., and Lackner, A.A. (1998). Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. Science *280*, 427-431.

Venkatesan, N.A.a.S. (1988). Nef Protein of HIV-1 is a Transcriptional Repressor of HIV-1 LTR. Science *241*, 1481-1485.

Vodyanik, M.A., Bork, J.A., Thomson, J.A., and Slukvin, II (2005). Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. Blood *105*, 617-626.

Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., and Alizon, M. (1985). Nucleotide sequence of the AIDS virus, LAV. Cell *40*, 9-17.

Walker, B.D. (2007). Elite control of HIV Infection: implications for vaccines and treatment. Topics in HIV medicine : a publication of the International AIDS Society, USA *15*, 134-136.

Wang, L., Mukherjee, S., Jia, F., Narayan, O., and Zhao, L.J. (1995). Interaction of virion protein Vpr of human immunodeficiency virus type 1 with cellular transcription factor Sp1 and trans-activation of viral long terminal repeat. The Journal of biological chemistry *270*, 25564-25569.

Westendorp MO, F.R., Ochsenbauer C, Stricker K, Dhein J, Walczak H, Debatin KM, Krammer PH. (1995). Sensitization of T cells to CD95-mediated apoptosis by HIV-1 Tat and gp120. Nature *375*, 497-500.

Wightman, F., Solomon, A., Khoury, G., Green, J.A., Gray, L., Gorry, P.R., Ho, Y.S., Saksena, N.K., Hoy, J., Crowe, S.M., *et al*. (2010). Both CD31(+) and CD31(-) naive CD4(+) T cells are persistent HIV type 1-infected reservoirs in individuals receiving antiretroviral therapy. The Journal of infectious diseases *202*, 1738-1748.

Winston, J.A., Bruggeman, L.A., Ross, M.D., Jacobson, J., Ross, L., D'Agati, V.D., Klotman, P.E., and Klotman, M.E. (2001). Nephropathy and establishment of a renal reservoir of HIV type 1 during primary infection. The New England journal of medicine *344*, 1979-1984.

Wolf D, G.S. (2007). TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. Cell *131*, 46-57.

Wolf D, G.S. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature *458*, 1201-1204.

Wolf D, H.K., Goff SP. (2008). TRIM28 mediates primer binding site-targeted silencing of Lys1,2 tRNA-utilizing retroviruses in embryonic cells. Proc Natl Acad Sci U S A *105*, 12521-12526.

Wonderlich, E.R., Leonard, J.A., Kulpa, D.A., Leopold, K.E., Norman, J.M., and Collins, K.L. (2011). ADP ribosylation factor 1 activity is required to recruit AP-1 to the major histocompatibility complex class I (MHC-I) cytoplasmic tail and disrupt MHC-I trafficking in HIV-1-infected primary T cells. Journal of virology *85*, 12216-12226.

Wong-Staal, F., Shaw, G.M., Hahn, B.H., Salahuddin, S.Z., Popovic, M., Markham, P., Redfield, R., and Gallo, R.C. (1985). Genomic diversity of human T-lymphotropic virus type III (HTLV-III). Science *229*, 759-762.

Wong, J.K., Hezareh, M., Gunthard, H.F., Havlir, D.V., Ignacio, C.C., Spina, C.A., and Richman, D.D. (1997). Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. Science *278*, 1291-1295.

Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. American journal of human genetics *89*, 82-93.

Wu, Q., Shi, H., Suo, Z., and Nesland, J.M. (2003). 5'-CpG island methylation of the FHIT gene is associated with reduced protein expression and higher clinical stage in cervical carcinomas. Ultrastructural pathology *27*, 417-422.

Wu S, M.H. (2011). Response-adaptive regression for longitudinal data. Biometrics *67*, 852-860.

Yabut, O., and Bernstein, H.S. (2011). The promise of human embryonic stem cells in aging-associated diseases. Aging *3*, 494-508.

Yamauchi M, F.B., Khan C, Berwin B, Barklis E. (1995). Stem cell factor binding to retrovirus primer binding site silencers. J Virol *69*, 1142-1149.

Yao, F. (2007). Functional principle component analysis for longitudinal and survival data. Statistica Sinica *17*, 965-983.

Yao F, M.H., Wang JL. (2005). Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association *100*, 577-590.

Yedavalli, V.S., Benkirane, M., and Jeang, K.T. (2003). Tat and trans-activation-responsive (TAR) RNA-independent induction of HIV-1 long terminal repeat by human and murine cyclin T1 requires Sp1. The Journal of biological chemistry *278*, 6404-6410.

Yeung, M.L., Houzet, L., Yedavalli, V.S., and Jeang, K.T. (2009). A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. The Journal of biological chemistry *284*, 19463-19473.

Ylisastigui L, C.J., Rucker VC, Melander C, Bosch RJ, Brodie SJ, Corey L, Sodora DL, Dervan PB, Margolis DM (2004). Polyamides reveal a role for repression in latency within resting T cells of HIV-infected donors. J Infect Dis *190*, 8.

Yu, G.R., Qin, W.W., Li, J.P., Hua, W., Meng, Y.L., Chen, R., Yan, B., Wang, L., Zhang, X., Jia, L.T., *et al*. (2012). HIV-TAT-fused FHIT protein functions as a potential pro-apoptotic molecule in hepatocellular carcinoma cells. Bioscience reports *32*, 271-279.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, II, and Thomson, J.A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. Science *324*, 797-801.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., *et al*. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Zablotska, I. (2013). Ending the pandemic: reducing new HIV infections to zero. Journal of the International AIDS Society *16*, 18933.

Zauli G, G.D., Secchiero P, Dutartre H, Olive D, Capitani S, Collette Y. (1999). Human immunodeficiency virus type 1 Nef protein sensitizes CD4(+) T lymphoid cells to apoptosis via functional upregulation of the CD95/CD95 ligand pathway. Blood *93*, 1000-1010.

Zaunders J, D.W., Churchill M. (2011). The Sydney Blood Bank Cohort: implications for viral fitness as a cause of elite control. Current opinion in HIV and AIDS *6*, 151-156.

Zhang, H., Zhou, Y., Alcock, C., Kiefer, T., Monie, D., Siliciano, J., Li, Q., Pham, P., Cofrancesco, J., Persaud, D., *et al*. (2004). Novel single-cell-level phenotypic assay for residual drug susceptibility and reduced replication capacity of drug-resistant human immunodeficiency virus type 1. Journal of virology *78*, 1718-1729.

Zhang, H.S., Ruan, Z., and Sang, W.W. (2011). HDAC1/NFkappaB pathway is involved in curcumin inhibiting of Tat-mediated long terminal repeat transactivation. Journal of cellular physiology *226*, 3385-3391.

Zhang, H.S., and Wu, M.R. (2009). SIRT1 regulates Tat-induced HIV-1 transactivation through activating AMP-activated protein kinase. Virus research *146*, 51-57.

Zhang J, S.D., Crumpacker CS. (2007). Primitive hematopoietic cells resist HIV-1 infection via p21. J Clin Invest *117*, 473-481.

Zhang, J.L., Sharma, P.L., and Crumpacker, C.S. (2000). Enhancement of the basal-level activity of HIV-1 long terminal repeat by HIV-1 nucleocapsid protein. Virology *268*, 251-263.

Zhang, L., Ramratnam, B., Tenner-Racz, K., He, Y., Vesanen, M., Lewin, S., Talal, A., Racz, P., Perelson, A.S., Korber, B.T., *et al*. (1999a). Quantifying residual HIV-1 replication in patients receiving combination antiretroviral therapy. The New England journal of medicine *340*, 1605-1613.

Zhang, Z., Ersoz, E., Lai, C.Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., *et al*. (2010). Mixed linear model approach adapted for genome-wide association studies. Nature genetics *42*, 355-360.

Zhang, Z., Schuler, T., Zupancic, M., Wietgrefe, S., Staskus, K.A., Reimann, K.A., Reinhart, T.A., Rogan, M., Cavert, W., Miller, C.J., *et al*. (1999b). Sexual transmission and propagation of SIV and HIV in resting and activated CD4+ T cells. Science *286*, 1353-1357.

Zhou M, H.M., Radonovich MF, Kashanchi F, Peng J, Price DH, Brady JN. (2000). Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription. Mol Cell Biol *20*, 5077-5086.

Zhu, H., Shyh-Chang, N., Segre, A.V., Shinoda, G., Shah, S.P., Einhorn, W.S., Takeuchi, A., Engreitz, J.M., Hagan, J.P., Kharas, M.G., *et al*. (2011). The Lin28/let-7 axis regulates glucose metabolism. Cell *147*, 81-94.

Zhu, T., Muthui, D., Holte, S., Nickle, D., Feng, F., Brodie, S., Hwangbo, Y., Mullins, J.I., and Corey, L. (2002). Evidence for human immunodeficiency virus type 1 replication *in vivo* in CD14(+) monocytes and its potential role as a source of virus in patients on highly active antiretroviral therapy. Journal of virology *76*, 707-716.