CELLULAR AND VIRAL FACTORS AFFECTING HIV-1 SILENCING AND REACTIVATION

Nicholas James Norton

Queens' College

July 2018

This dissertation is submitted for the degree of Doctor of Philosophy

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

Table of Contents

Acknowledgements

I would like to thank Professor Andrew Lever for the opportunity to work on this project and for his thoughtful guidance throughout. I am grateful to many members of the Lever group but in particular to Dr Hoi Ping Mok, Dr Axel Fun and Dr Jing Garland who assisted me in finding my feet in the laboratory and in learning many of the techniques in this project.

I would like to gratefully acknowledge the insight and guidance of Professor John Sinclair and Dr Mark Wills given over many cups of coffee. I am also grateful to Professor Paul Lehner, Dr Nick Matheson and to Dr Iva Tschasovnikarova for the materials and assistance they have provided to this project. I am obliged to Jack Hirst for technical assistance in the lab as well as to Dr Zoya Butt and Dr Fatima Sharif, two students from Aga Khan University, who carried out experiments under my guidance.

Finally I give thanks to my partner Richard for all the support he has shown me.

This project is supported by the Medical Research Council and the British Infection Association

Summary

Despite advances in the treatment of HIV-1 a cure remains elusive. A significant barrier to the eradication of the virus from an infected individual is a pool of cells infected with transcriptionally silent proviruses. A key pillar of the strategy to eradicate latent viruses has been called 'kick and kill', whereby the latent virus is stimulated to transcribe rendering the host cell vulnerable to eradication by cytotoxic T cells. Optimising the reactivation signal is therefore critical to this approach. Here the established model system of latency 'J-lat' is used to probe optimum reactivation signals. Single clones are observed to respond to maximal stimulation with a single agent with a fixed proportion of cells. Here it is shown that this proportion can be overcome by dosing with two agents in combination and critically that maximum synergies between agents occur at concentrations of agents close to those achieved *in vivo*. The role of SETDB1 recruitment by the recently described HUSH complex is examined using shRNA knockdowns of these proteins. Knockdown does not increase expression from the majority of J-lat clones tested. Viral factors which influence silencing and reactivation from latency have not been explored to the same extent. Here mutations affecting the binding of splicing factors to HIV-1 mRNA were cloned into laboratory viruses. A reduction in splice factor binding is seen to change the use of splice junctions required for the production of Tat mRNA; in turn this alters the rate at which proviruses are silenced. In addition the threshold for transcription in response to stimulation is increased in mutants with reduced splice factor binding.

Chapter 1: Introduction

1.1 HIV historical context and epidemiology

The virus now known as the Human Immunodeficiency Virus (HIV) was first isolated in the laboratory of Luc Montagnier(Barré-Sinoussi *et al.*, 1983)in 1983. The virus was identified from patients with Acquired Immunodeficiency Syndrome (AIDS). The first sign of this emerging epidemic was a 1981 report in the Centers for Disease Control's *Morbidity and Mortality Weekly Report* of 5 cases of pneumonia caused by the uncommon pathogen *Pneumocystis carinii* (now *P. jirovecii*) amongst homosexual men(MMWR, 1982a). This was soon followed by reports of Kaposi's sarcoma(MMWR, 1982a) and persistent generalised lymphadenopathy(MMWR, 1982b) also occurring exclusively among homosexual men. A report in July of 1982 of AIDS occurring in patients with Haemophilia suggested that the aetiologic agent was transmissible by blood products(MMWR, 1982c). By 1985 the causal link between AIDS and the virus identified by Montagnier had been almost universally accepted(Marx, 1985).

AIDS was reported in South Africa in 1982 among gay men; not long after this an epidemic affecting the general population in sub Saharan Africa was noted. Recent figures from the World Health Organisation estimate the number of people alive with HIV infection in 2014 was 36.9 million and that 2 million new infections had occurred in that year(WHO, 2015). The majority (25.8 million people) live in sub-Saharan Africa. Since the identification of the virus and description of its genome it has become clear from detailed phylogenetic studies that the global epidemic has its origins in Africa(Sharp *et al.*, 2001; Worobey *et al.*, 2008)

Two distinct HIV types exist; HIV-1 and HIV-2. HIV-2 is primarily confined to west Africa(Campbell-Yesufu and Gandhi, 2011) and is less pathogenic than HIV-1. Genetic analysis of HIV-1 reveals three groups; M, N and O. Group M is the most common accounting for 90% of cases; groups N and O are found primarily in Cameroon (Gifford *et al.*, 2007). Group M has been further divided into subtypes

which show variations in their geographic distribution(Sharp *et al.*, 2001). Ape to human transmission of the Simian Immunodeficiency Viruses affecting the chimpanzee (*Pan troglodytes troglodytes*) and Gorilla (*Gorilla gorilla gorilla*) is thought to be the origin of HIV-1(Sharp *et al.*, 2001; Keele *et al.*, 2006; Van Heuverswyn *et al.*, 2006). A number of analyses have estimated the date of this cross species transmission to the early 20th Century (Worobey *et al.*, 2008; Sharp and Hahn, 2011; Faria *et al.*, 2014).

The development of treatment for HIV infection began with the introduction of azidothymidine (zidovudine, AZT) in 1987. Further drug discoveries followed, however their effect on the virus was only temporary and patients' disease progressed. Truly effective therapy began in the 1990's following the publication of the finding that combining available drugs to create highly active anti-retroviral therapy (HAART) led to vastly better outcomes(Gulick *et al.*, 1997; Hammer *et al.*, 1997).

Developing a vaccine to prevent the spread of HIV is an attractive goal but one that has remained elusive. The nature of the viral life cycle ensures that viral proteins are frequently mutated meaning that immune responses targeting particular viral epitopes are unlikely to be able to achieve the necessary breadth to cover all variations. Furthermore as immune clearance of the virus does not occur in treated or untreated infection there has been no clear indicator of what qualitative or quantitative immune response is needed to guide generation of a successful vaccine(Fauci and Marston, 2015). Early vaccine efforts yielded promising antibody titres but these antibodies were found not to neutralise(Dolin *et al.*, 1991). The more recent finding of the occurrence of broadly neutralising antibodies in those who have been infected for a long time has revealed new targets for vaccine design (Haynes *et al.*, 2012). While these broadly neutralising antibodies have been shown to be protective when used by passive transfer in animal models (Shingai *et al.*, 2014) they have so far not been induced in humans by vaccination (Dosenovic *et al.*, 2015). In addition to the desire to develop a vaccine for primary prevention, research is also on-going into how vaccines may

produce augment strategies to cure HIV by providing a boost to existing immune responses.

1.2 Clinical History of HIV-1 infection

HIV transmission occurs by the transfer of virus-infected bodily fluids from one person to another. The most common mechanisms of transmission are unprotected sexual intercourse, needle sharing and the administration of contaminated blood products. Although a proportion of people are symptomatic during primary HIV infection the symptoms are non specific and therefore frequently missed. Symptoms occur around 10 days post infection and include fever, sore throat, myalgia, lymphadenopathy and a maculopapular rash. Circulating levels of virus reach a peak during the initial illness before falling to a set point. Similarly numbers of CD4+ lymphocytes fall during primary infection before rebounding. After the primary phase of infection patients then enter 'clinical latency'. This stage may last many years, during which the patient is commonly asymptomatic. Nonetheless viral replication is ongoing and detectable levels of virus will be present in the patient's plasma. There is a slow fall in CD4 count during this period.

Eventually the CD4 count will decline to the point that the immune system is significantly compromised and opportunistic infections and/or HIV associated malignancies start to occur. These often herald a loss of immune control of HIV and a rise in the patient's viral load. According to the Centers for Disease Control (CDC) criteria patients are deemed to have AIDS when the CD4 count reaches 200 cells/mm^3 or if any of a number of AIDS defining conditions occurs. In one US study of patients before the advent of effective antiretroviral therapy the survival from diagnosis of a first AIDS defining condition was 3-51 months(Gail *et al.*, 1997).

In a minority of patients the clinical syndrome does not progress as described above. In approximately 5% of infected patients the CD4 count remains stable or falls at a much slower rate despite the presence of detectable viraemia meaning

that these patients progress to AIDS at a much later time point. This group of patients are often referred to as Long Term Non Progressors. The factors which give rise to this altered phenotype of infection are not fully understood however an enhanced CD8 response has been identified in a number of studies(Sheppard *et al.*, 1993; Keet *et al.*, 1994; Cao *et al.*, 1995). Enhanced CD8 responses have also been linked to the phenomenon of highly exposed persistently seronegative individuals. This group was initially described among female sex workers in Kenya and Gambia and have been found to have specific CD8+ T cell responses to HIV peptides, suggesting prior exposure to HIV, but do not have detectable plasma HIV RNA or antibodies to HIV proteins(Rowland-Jones *et al.*, 1995; Fowke *et al.*, 1996; Kaul *et al.*, 2001).

Similarly there exists a yet smaller subpopulation of patients who are able to suppress the virus to an undetectable level without the need for antiretroviral therapy(reviewed in Saag & Deeks 2010). Like the non-progressors the exact mechanisms responsible for this phenotype are not understood. As the virus found in these patients has been found to replicate *ex vivo* (Blankson *et al.*, 2007)it is believed that the mechanisms of control are immunological.

In contrast, 15-20% of children experience a rapidly progressive disease course. Untreated these patients progress to severe immunodeficiency within the first year. Although not fully understood it has been found that polymorphisms in the co-receptor CCR5 are associated with this phenotype(Smith *et al.*, 1997; Singh and Spector, 2009).

The observation that not all those infected respond to infection in the same way has influenced approaches to HIV cure. If the beneficial phenotypes above could be induced in an infected patient it would lead to a 'functional cure' i.e. the patient would not need to take antiretroviral therapy to prevent progression of their HIV. This is discussed further below.

Class of antiretroviral	Examples	Mechanism of action
Nucleoside/nucleotide analogue reverse transcriptase inhibitor (NNRTI)	Abacavir Lamivudine Emtricitabine Tenofovir	Are added to nascent cDNA by HIV reverse transcripase and cause chain termination
Non-nucleoside reverse transciptase inhibitor (NRTI)	Efavirenz Niviripine Rilpivirine	Allosterically bind to reverse transcriptase and cause conformational change which disables the active site
Integrase strand transfer inhibitor (INSTI)	Raltegravir Dolutegravir Elvitegravir	Inhibit the joining of HIV proviral DNA into cellular DNA by binding magnesium ions at the active site
Protease inhibitor (PI)	Ritonavir Lopinavir Atazanavir Darunavir	Inhibit proteolytic cleavage of gag/pol polyproteins preventing maturation of the virion into and infectious particle
Entry inhibitor	Maraviroc	Binds to CCR5 preventing HIV envelope glycoprotein from binding
Fusion inhibitor	Enfuvirtide	Binds to envelope protein gp41 and prevents a conformational change essential for fusion of the virion to the cellular membrane

Table 1.1 – Classes of antiretroviral drugs and their mechanism of action. The examples given represent agents currently in clinical use in the UK. Most antiretroviral regimens consist of two agents from the NRTI class plus one agent from another class.

Treatment for HIV consists of combinations of drugs. Normally this will include three or more drugs from at least two different classes(Table1.1). By targeting multiple steps in the virus life cycle combination therapy is not only able to be highly effective in reducing viral replication but is also effective in preventing resistance from emerging. National and international guidelines differ on the timing of the initiation of antiretroviral therapy. Early treatment in asymptomatic patients with good CD4 counts risks patients developing side effects from the treatment at a time when they are at low risk of opportunistic infections. Side effects of antiretroviral therapy differ according to the exact composition of the regimen but include increased risk of cardiovascular disease, renal toxicity and neuropsychiatric disturbance. In addition a number of antiretroviral drugs have significant effects on the metabolism of other drugs. Once initiated the aim of therapy is to reduce the number of circulating copies of viral RNA to below the limit of detection of a standard clinical assay (approximately 50 copies/ml). Patients on therapy with fully suppressed viral loads do not progress to AIDS. In addition starting antiretroviral therapy has been shown to reduce the risk of serious non AIDS events regardless of the CD4 count at the time of initiating treatment(INSIGHT START 2015). Reducing the viral load also has benefits in terms of reducing the risk of transmission to others both mother to child(Townsend *et al.*, 2008) and in serodiscordant couples(Cohen *et al.*, 2011).

1.3 HIV-1 life cycle

1.3.1 Overview

HIV-1 carries two copies of its positive sense RNA genome. Upon entry into a host cell the RNA genome is converted into DNA by the viral enzyme reverse transcriptase. The viral genomic DNA interacts with viral proteins to produce a pre-integration complex; this enters the nucleus of the cell via the nuclear pore. The viral enzyme integrase breaks the cellular genomic DNA and inserts the viral DNA at which point it is termed a provirus. At each end of the viral DNA is an identical sequence known as a long terminal repeat (LTR). Once inserted the HIV genome functions in a similar way to a cellular gene with the 5' LTR acting as a

promoter. The provirus encodes a single mRNA capable of producing a number of gene products through alternative splicing. The initial gene products are multiply spliced mRNAs encoding small regulatory HIV proteins. Later gene products encode proteins required for the formation and maturation of the capsid, the envelope glycoproteins responsible for cell tropism and entry and the enzymes required for reverse transcription and integration. To complete its life cycle two copies of the unspliced genomic RNA are packaged by the group specific antigen (Gag) polyprotein through its nucleocapsid domain. The polyprotein is cleaved during viral assembly and budding with the p24 domain forming the viral capsid. At the same time the envelope glycoprotein (Env) assembles on the cell surface. The viral capsid buds off the cell taking with it a cell membrane derived envelope containing viral envelope glycoproteins. The sections below detail the parts of the life cycle explored in this project.

1.3.2 Cell entry and cell tropism

HIV envelope glycoprotein binds to CD4 on the cell membrane of a target cell. This binding alone is not sufficient to allow HIV entry to a cell; the protein must also bind to a chemokine co-receptor. As gp120 binds to CD4 a structural change occurs which exposes a co-receptor binding site. *In vivo* the co receptor molecules are CXCR4 and CCR5(Clapham, 2001); the requirement for the virus to bind either CCR5 or CXCR4 is referred to as its tropism. Host cells with these surface proteins include T-lymphocytes, monocytes and dendritic cells. Binding of the co-receptor molecule initiates a further structural change to the envelope glycoprotein which triggers fusion of the viral envelope with the cell membrane.

Genetic reconstructions have shown that infections start with a single virus and in the overwhelming majority of cases that virus was CCR5 tropic(Salazar-Gonzalez *et al.*, 2009). This means that early in infection virions requiring the CCR5 co-receptor binding predominate. Individuals who are homozygous for a mutation in CCR5 that prevents it localising to the cell surface have marked resistance to infection with HIV (Wilkinson *et al.*, 1998; Philpott *et al.*, 1999) suggesting that CCR5 plays a significant role in establishing HIV infection. Inhibition of the

binding of HIV to CCR5 is the mechanism by which the HIV entry inhibitor maraviroc has its effect.

CXCR4 tropic viruses appear to arise later in infection. The factors that cause the switch from CCR5 to CXCR4 tropism are not understood. Due to the distribution of CXCR4 across cell types viruses exploiting it as a co-receptor have a broader range of target cells available to them; on T cells the expression of CCR5 is limited to memory cells whereas CXCR4 is expressed by a number of T cell subsets including naïve cells. Patients with CXCR4 tropic virus have been found to harbour proviral DNA in a higher proportion of their naïve T cells(Ostrowski *et al.*, 1999).

1.3.3 Reverse transcription

Once inside the host cell the dimeric RNA genome of HIV must be reverse transcribed to make double stranded DNA(For a review see Hu & Hughes 2012). HIV reverse transcriptase (RT) requires primer binding to the RNA to initiate transcription; *in vivo* tRNA(Lys3)binds to the primer binding site 180nt from the 5' end of the genomic RNA. Reverse transcription proceeds to the 5' terminal R region of the genomic RNA, the RNase activity of RT degrades the R/U5 template RNA. The complementary DNA then binds to the identical R region at the 3' end of the genomic RNA and reverse transcription proceeds to the original primer binding site and the RNA template is degraded. The polypurine tract is resistant to the RNase activity of RT and primes positive strand DNA synthesis which proceeds to the end of the tRNA primer attached to the negative strand DNA, this results in two DNA strands which are complementary at the primer binding site. Finally the two DNA strands bind at the primer binding site and extension of both strands is completed resulting in double stranded genomic DNA with a long terminal repeat (LTR) at each end.

As HIV RT lacks proof reading ability the process of conversion from RNA to DNA is error prone. Early studies using purified enzyme *in vitro* reported very high error rates in the region of 1 per 1000 to 1 per 5000 bases(Preston, Poiesz and Loeb, 1988; Roberts, Bebenek and Kunkel, 1988) and suggested that the genetic

diversity of HIV observed *in vivo* was due largely to RT. Studies which have sequenced viruses produced in more complete infection models have yielded a lower estimate of the error rate at $3.4x10^{-5}$ per bp per replication cycle(Mansky and Temin, 1995; Gao *et al.*, 2004). This lower estimate represents the combined errors introduced by both RT and cellular RNA Polymerase II and suggests that the presence of other HIV proteins *in vivo* improves the fidelity of RT. Nonetheless the per cycle mutation rate of HIV means that a large sequence diversity is seen soon after infection. This is a reflection of the sheer number of replication events occurring during early infection.

1.3.4 Integration

Integration of HIV proviral DNA into host genomic DNA is mediated by the viral protein integrase (IN). This protein is packaged along with reverse transcriptase when the viral particle is made. After reverse transcription double stranded viral DNA remains closely associated to IN in the pre-integration complex. Viral DNA is prepared for integration by the removal of the two terminal nucleotides from each 3' end leaving a terminal CA-3' dinucleotide. Each end attaches to host genomic DNA 5 nucleotides apart on opposite strands. The resulting single stranded gaps are filled in by cellular enzymes leaving a duplicated sequence at each end of the provirus (Craigie and Bushman, 2012).

Binding of HIV integrase to the cellular protein LEDGF (reviewed in Christ & Debyser 2013) is essential for integration to take place. LEDGF tethers the integrase to chromatin and stabilises it. This interaction likely has a significant role in determining the site of integration into the host genome.

The selection of integration site in HIV has been demonstrated to be only weakly sequence dependent(Holman and Coffin, 2005; Wu *et al.*, 2005). One study which utilised a large data set of over 40000 integration sites mapped by sequencing identified non sequence determinants of integration sites(Wang *et al.*, 2007). They showed that HIV has a preference for integration in the major groove of chromatinised DNA. They also observed that local histone marks which associate with permissive transcription were associated with more frequent integration

events, similarly sites associated with the repressive chromatin mark H3K27me3 were negatively associated with integration events. The authors also studied DNA methylation and showed that HIV integration was disfavoured by areas of DNA methylation.

It has been observed that HIV preferentially integrates into actively transcribed cellular genes (Schröder *et al.*, 2002; Brady *et al.*, 2009). Microscopy work done the laboratory of Marina Lusic has found that integration occurs preferentially at the nuclear periphery close to nuclear pores (Marini *et al.*, 2015). As actively transcribed genes will localise to the nuclear pore to facilitate efficient export of mRNA these findings together suggest that HIV integrates into those parts of DNA closest to its point of entry into the nucleus.

Failure to complete integration can result in the creation of circularised forms of HIV proviral DNA. The longevity of these forms of DNA is not fully established. It has been demonstrated that transcription can occur from unintegrated DNA(Chan *et al.*, 2016). Furthermore it has been observed that IN is capable of cleaving circularised DNA and completing integration(Thierry *et al.*, 2015).

1.3.5 Transcriptional control

The 5' LTR functions as a promoter for integrated proviral DNA and will recruit RNA polymerase II to initiate transcription. Pol II requires phosphorylation by CDK9 in order for productive elongation to occur. CDK9 together with cyclinT1 make up the positive transcriptional elongation factor b (P-TEFb) which is recruited to Pol II by the viral accessory protein Tat (transactivator of transcription). The initiated but stalled RNA transcripts contain a highly conserved RNA secondary structure the Transactivation Response element(TAR) to which Tat binds to mediate this process. In addition to P-TEFb Tat recruits other positive signals for transcriptional elongation, proteomic studies have identified as binding partners for Tat the transcription factors AFF4, ENL, AF9, and the elongation factor ELL2(He *et al.*, 2010).

Once productive transcription occurs more Tat protein is translated locking transcription into a positive feedback loop. Free Tat in the nucleus is able to release further P-TEFb from where it is sequestered by the protein HEXIM1 in the 7SK RNP. Tat is essential to productive transcription, studies of systems in which Tat has been removed show the accumulation of abortive short transcripts stalled at the stage of TAR loop production; a transcript 59 bases long; but without elongation.

Post-translational modification of Tat also has a role to play in regulating transcription. Acetylation of Tat by the cellular protein p300 occurring at lysine residues in the RNA binding site of Tat has been found to be essential for efficient elongation of transcription(Ott *et al.*, 1999; Deng *et al.*, 2000). Conversely methylation of Tat lysine residues by cellular methyltransferases has been linked to reduced efficiency of transactivation(Van Duyne *et al.*, 2008).

Modifications to the local histone environment necessary for HIV transcription and a more detailed description of transcriptional control are discussed below.

1.3.6 HIV-1 RNA splicing

Despite a relatively short genome of 9.7kb HIV is able to produce 9 different proteins and polyproteins from a single transcription start site. Alternative splicing allows HIV to produce over 50 mRNA species by utilising 4 splice donors and 7 splice acceptors (Ocwieja *et al.*, 2012; Emery *et al.*, 2017). These products are grouped by size into three classes, 1.8kb completely spliced, 4kb incompletely spliced, and unspliced transcripts. In general the completely spliced transcripts code for the small accessory proteins, Tat, Rev, Vif and Nef whereas the incompletely spliced transcripts code for Env, Vpr and Vpu. Unspliced transcripts are translated to yield Gag and Gag/Pol in addition to being packaged as genomic RNA.

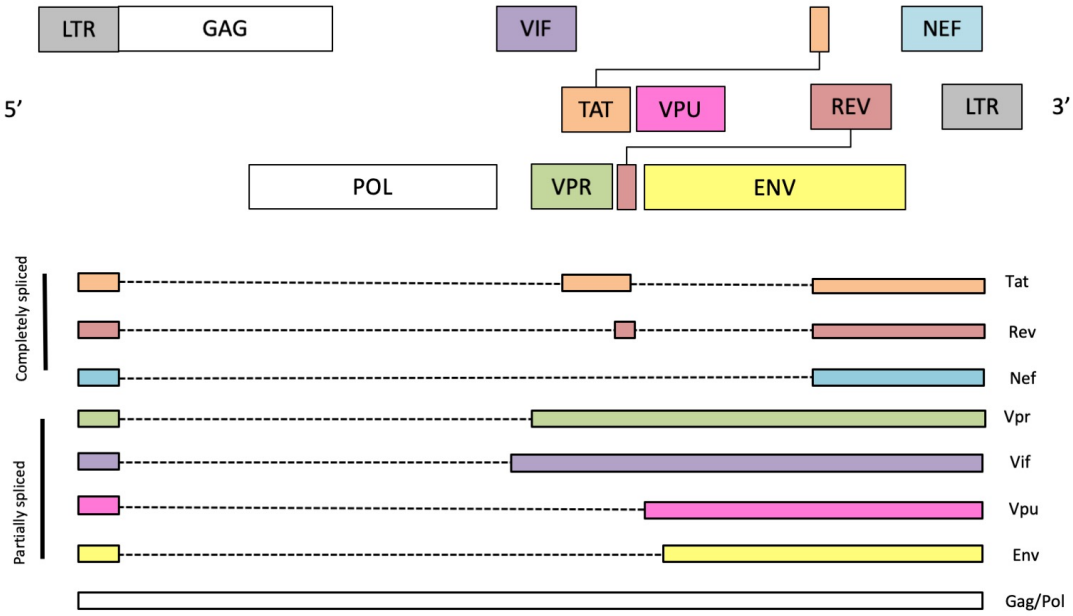


Figure 1.1 – Schematic representation of the generation of HIV transcripts by alternative splicing. The top half of the figure shows the arrangement of the HIV genome demonstrating the proteins and their organisation in three reference frames. The bottom half shows how the RNA transcribed from the 5' LTR is spliced to yield different protein products. The 3' end of the RNAs are polyadenylated (not shown).

NB the gene products are scaled to show their relative organisation and are not drawn to scale with respect to the size of the coding region

Figure 1.1 shows a schematic of the major mRNA species and their formation from combinations of 5' splice donors and 3' acceptors. The splice site usage is regulated by a number of *cis*-acting elements (Reviewed in Karn & Stoltzfus 2012; Martin Stoltzfus 2009). Exonic splice enhancers (ESEs) are sequences within exons which are identified and bound by cellular serine-arginine rich proteins

Completely spliced HIV mRNAs exit the nucleus by the same mechanisms as cellular mRNAs. Both unspliced and incompletely spliced mRNA species contain the Rev Response Element (RRE) a structured 351 nucleotide element which binds the viral Rev protein to facilitate the export of the RNA. In the absence of Rev these transcripts would be retained in the nucleus and degraded. The early phase of HIV transcription is characterised by the production of completely spliced mRNAs and their protein products, including Rev; late phase transcription leads to the production of mRNAs required for virion structure and assembly. The accumulation of Rev is therefore a determinant of the switch from early to late expression profiles.

1.4 HIV-1 latency

As noted above treatment for HIV can achieve long term virological suppression however on stopping therapy infection is rekindled. This rebound of virus production likely originates from inactive or slowly replicating pools of 'latent' virus that subsequently escape to infect more permissive cells, amplifying the virus. Early evidence for proviral latency in HIV came from experiments in which a stably transduced cell line could be induced to make virions but was quiescent without activation(Folks *et al.*, 1986). This suggested that, despite carrying an integrated provirus, production of virions was stalled at a point downstream of transcription. As many of the activators of HIV transcription are inducible on activation of a CD4+ T cell it appeared likely that latency would occur more frequently in resting T cells. These cells have been shown to present barriers to productive infection(Zhou *et al.*, 2005) in particular the presence of SAMHD1 in resting cells reduces the efficiency of reverse transcription(Zack *et al.*, 1992). Studies of primary resting CD4+ T cells isolated from HIV infected donors have

shown that they can be induced to make replication competent virus(Chun et al. 1997; Han et al. 2004; Crooks et al. 2015).

The apparent difficulty in infecting resting CD4+ T cells *ex vivo* has meant that the mechanism by which these cells become infected *in vivo* is not clear. A widely accepted explanation is that the CD4+ T cells are infected while in an activated state. Productive infection with HIV will cause many of these cells to die. A small number of CD4+ cells will return to a resting state as a memory T cell. If an infected cell survives to become a memory cell the change in the cellular environment leads to a reduction in the transcription factors available to support HIV transcription and virus production. The provirus therefore becomes silent until the cell is activated. However work done in the laboratory of Eric Verdin has challenged this and shown that it is possible to directly establish a latent infection in both activated and resting CD4+ T cells(Chavez, Calvanese and Verdin, 2015).

The long life span of resting CD4+ cells means that, once infected with HIV, they are able to form a reservoir of latent virus which, in treated individuals, decays slowly over the course of infection(Siliciano and Siliciano, 2015). Their longevity also helps to explain the observation that drug resistant viruses can be archived for many years after switching antiretroviral therapy(Noë, Plum and Verhofstede, 2005). Despite the longevity of infected resting CD4+ T cells they are rare; in an infected individual the frequency of latently infected cells from which replication competent virus can be recovered is approximately 1 per 10^6 resting CD4+ cells(Chun et al. 1997).

1.4.1 Clonal expansion of latently infected cells

One mechanism that has been proposed for the maintenance of the latent viral reservoir is clonal expansion of latently infected cells. Homeostatic proliferation of T cells maintains the lymphocyte population, In this model a latently infected cell undergoes homeostatic proliferation in so doing any integrated proviral DNA is copied along with the host cell DNA and passed to the progeny cells. As cell line models of HIV latency show that cell cycling does not disrupt latency the latent

phenotype may be passed on to the daughter cells. Studies have identified that clonal proviral sequences appear in the circulating lymphocytes of patients on treatment(Chomont *et al.*, 2009; Wagner *et al.*, 2013) and that infection can occur in the progenitor cells which later expand to form mature T cell populations(Buzon *et al.*, 2014).

The expansion and persistence of certain clones may be driven by integration of HIV DNA into genes responsible for controlling the cell cycle. One study using cells from patients on ART has identified a predilection for integration into the BACH2 and MLK (Maldarelli *et al.*, 2014) genes which the authors point out are associated with cell cycling and in which mutations have been identified in some tumours. They hypothesise that the observed excess of cells with integration events in these genes is due to an enhanced survival conferred by the integration of an HIV provirus and subsequent alteration of the transcriptional profile of the gene. A similar study carried out by sampling three patients over 12 years also demonstrated clonally expanded integrations in BACH2 as well as integrations into MDC1 and IZKF3 which are both associated with oncogenesis(Wagner *et al.*, 2014). In contrast a large-scale study of over 6000 integration sites recovered from 11 patients found that while there was an enrichment of integrations into cancer genes compared to all genes there was no enrichment when comparing cancer genes to transcriptionally active genes(Cohn *et al.*, 2015).

The extent to which the proviruses in clonally expanded integration sites are replication competent has been a subject of intense discussion. Recently it has been demonstrated that the viraemia induced by treating infected individuals with romidepsin, an agent designed to induce transcription from latent viruses, contains multiple identical sequences (Winckelmann *et al.*, 2017). Furthermore studies which have isolated resting CD4 cells from infected patients and then sequenced the viruses which emerge from the cells have identified identical viral sequences (Bui *et al.*, 2017; Mok *et al.*, 2018). The enormous sequence diversity found in acute infection and occurring during the infection process through RT errors means that two viruses which emerge from different cells but have the same sequence can be assumed to represent clonal expansion of an integration

site rather than two independent integrations with identical proviruses. This work has also been bolstered by the finding that cells with replication competent HIV proviruses can undergo clonal expansion *ex vivo* when stimulated with PHA(Hosmane *et al.*, 2017).

1.4.2 Tissue reservoirs of virus

HIV distributes itself widely throughout the host and can be found in a number of different tissue types. The varying degree to which the antiretroviral activity of drugs and the immune system are able to penetrate these sites has lead to the theory that certain sites may provide sanctuaries that support ongoing viral replication despite therapeutic plasma drug levels. One such site is the central nervous system and HIV can be detected in cerebrospinal fluid and in the choroid plexus of infected individuals(Falangola *et al.*, 1995). Drug penetration the central nervous system across the blood-brain barrier is poor meaning that these viruses may be exposed to sub-optimal antiretroviral therapy.

The gastrointestinal mucosa is also considered to represent a different compartment for HIV replication. Viruses recovered from here have phenotypic(Barnett *et al.*, 1991) and genetic (Goudsmit *et al.*, 1998)differences from those found in circulating lymphocytes or in semen. A recent exploration of viral sequences from lymphoid tissue from patients on antiretroviral therapy has suggested that ongoing evolution is occurring (Lorenzo-Redondo *et al.*, 2016). In their paper the authors sampled inguinal lymph nodes, plasma virus and circulating PBMC from three HIV patients upon initiation of antiretroviral therapy and at three and six months. They observed evolution in the lymph nodes and evidence that virus from the lymph nodes was trafficking to the blood compartment. This observation is concerning as it suggests that current antiretroviral treatment is not sufficient to control HIV replication at all anatomical sites. A caveat to the author's conclusions was that they were unable to detect any drug resistant populations from the lymph nodes. The mechanism proposed for this was that the concentrations of antiretrovirals were low in the lymph nodes and at these low concentrations the fitness cost of a resistance

mutation means that wild type viruses were able to out compete the resistant virus.

The extent to which this compartmentalisation of HIV prevents its eradication is not yet fully understood however some clinical trials of HIV cure strategies are including HIV load from rectal mucosal biopsies as outcome measures in an effort to capture data on these reservoirs.

1.5 Transcriptional block in HIV-1 latency

The transition of an infected cell from an activated to a resting phenotype, and the resulting change in available transcription factors, offers a plausible explanation of how the initial silencing of HIV might occur. The means by which latency is maintained are not fully understood however a number of cellular processes have been identified as being involved. Early cell line work indicated that proviral genomes in latently infected cells were found in areas of dense heterochromatin but nonetheless could be induced into expression (Jordan, Defechereux and Verdin, 2001; Jordan, Bisgrove and Verdin, 2003). This observation agreed with our understanding of the function of heterochromatin in the control of cellular genes. Heterochromatin is a region in which DNA is tightly wound around histone proteins and made highly condensed. These dense regions are inaccessible to transcription factors and thus act as repressors of gene function. Gene expression on the other hand takes place in areas of relaxed DNA termed euchromatin (reviewed in Grewal & Jia 2007). Many studies of HIV silencing have identified chromatin marks as part of the mechanism maintaining silencing; these are discussed in detail below however they may not offer a complete explanation of HIV latency.

1.5.1 Chromatin and regulation of eukaryotic transcription

The central element of the transcriptional machinery producing mRNA in eukaryotes is RNA polymerase II (Pol II). In order to initiate transcription Pol II must be recruited to the core promoter; Pol II alone cannot recognise and bind to a promoter. This recruitment occurs through the general transcription factors (GTFs) referred to as TFIIA, -B, -D, -F, -G and –H. The transcription cycle starts

with the binding of the TATA box binding protein (TBP) subunit of TFIID to the TATA box (Matsui *et al.*, 1980). The TATA box is a region of the promoter containing the sequence 5'-TATAAA-3' where the weaker adenine-thymine hydrogen bonding permits unwinding of the double helix. TFIIA and TFIIB subsequently bind up- and downstream respectively and together permit binding of Pol II.

The remaining GTFs assemble with Pol II on the promoter and melt a short section of DNA enabling positioning of the template strand into the cleft of Pol II (Kostrewa *et al.*, 2009). At the same time a variety of transcriptional cofactors bind such as the multi-protein complex Mediator. This assembly of Pol II the GTFs and Mediator has been termed the pre-initiation complex and together has a mass of over 3000kD(Hahn, 2004). The final event in the initiation of transcription is the phosphorylation of the carboxy-terminal domain of Pol II which triggers the release of Pol II from the GTFs and begins the recruitment of factors required for productive elongation(Buratowski, 2003).

1.5.2 Structure of the nucleosome

The composition of the nucleosome was first elucidated in 1974 by Roger Kornberg (Kornberg, 1974; Kornberg and Thomas, 1974). It was later demonstrated by ligating nucleosome bearing genetic material to various promoters that nucleosomes prevented transcriptional elongation (Lorch, LaPointe and Kornberg, 1987). The nucleosome core consists of an octamer of histone proteins with two each of histone H2A, H2B, H3 and H4 (Luger *et al.*, 1997). These form a barrel shape around which 146bp of DNA is wrapped in 1.65 turns, histone protein H1 binds the DNA at either end of the coil to stabilise it. The binding of DNA by histone proteins is extremely stable in normal cellular conditions. Nucleosomes form throughout the genome and are the principal subunit of chromatin. Areas of densely associated nucleosomes and structural proteins make up heterochromatin. This highly condensed DNA-protein complex is associated with repression of genes.

Each histone protein consists of a central alpha helical region flanked by N and C terminal extensions. The central region is made up of three alpha helices with intervening loops. The longer central helices pair to form heterodimers of histone proteins H3 with H4 and H2A with H2B. The smaller alpha helices facilitate the binding together of heterodimers to form an octamer as well as providing a positively charged surface for DNA binding. The N terminal extensions of each of the histone subunits in addition to two C terminal tails supplied by H2A protrude from the nucleosome where they are available for posttranslational modification.

1.5.3 Histone modifications and modifying enzymes

The modification of the tails of histone proteins is intimately linked to transcriptional events with different modifications giving rise to either repression or activation of transcription. The term 'histone code' was popularised by Strahl and Allis who proposed that the highly specific nature of these modifications could be read as a language encoding the functional state of the gene (Strahl and Allis, 2000). Histone proteins may be modified by methylation, phosphorylation, ubiquitination and acetylation and at a number of different residues resulting in a large repertoire of histone marks. Table 1.2 shows a selection of the histone marks which have been described and their observed effect on transcription (for an in depth review see Kouzarides 2007).

Position	Modification	Effect on transcription
H3K4	Methylation	Activation
H3K9	Trimethylation	Repression
H3K27	Methylation	Repression
H4K20	Methylation	Repression
H3K56	Acetylation	Activation
H4K16	Acetylation	Activation
H3R2	Arginine methylation	Activation
H4R3	Arginine methylation	Activation
H2BK120	Ubiquitination	Activation
H2AK119	Ubiquitination	Repression
H3S10	Phosphorylation	Activation

Table 1.2 – Selected covalent histone marks and their effect on transcription. Modified from Li et al 2007.

The most studied modifications have been the acetylation and methylation of histone lysine residues. In general lysine acetylation is associated with transcriptional activation, lysine methylation on the other hand may encode both activation and repression depending on the particular residue targeted and the number of methyl groups attached to the residue. It has been postulated that histone modifications such as acetylation change the electrostatic properties of the affected residue and therefore cause loosening of the chromatin structure, permitting transcription. Unmodified histone tails are positively charged and therefore interact with the negatively charged phosphate backbone of DNA, the addition of acetyl groups neutralises the charge and therefore decreases the capacity of the histone to bind DNA(Struhl, 1998). In contrast the addition of methyl groups to lysine residues does not result in any loss of charge. Early studies using radiolabelled methyl groups gave rise to the belief that histone methylation is irreversible (Byvoet *et al.*, 1972) however this has been shown not to be the case (reviewed in Bannister & Kouzarides 2005).

The deposition of post-translational modifications on histone proteins is catalysed by a number of cellular enzymes. Three groups of enzymes have been described which catalyse the deposition of methyl groups onto histones. The SET domain containing proteins and the DOT1-like proteins are specific histone lysine methyltransferases, the protein arginine N-methyltransferase family are responsible for the methylation of histone arginines (Jenuwein *et al.*, 2000; Feng *et al.*, 2002). Within each family specific enzymes catalyse reactions at different residues. In the case of the SET family the active site consists of the central SET domain, the specificity of the domain is determined by sequences in the flanking regions.

In addition to enzymes controlling the deposition of methyl groups onto histone proteins a diverse group of enzymes exist to catalyse the addition and removal of acetyl groups. Histone acetylation typically occurs at multiple residues and is carried out by histone acetyl transferase (HAT) complexes(Brown *et al.*, 2000). The removal of histone acetyl groups is carried out by a group of enzymes referred

to as histone deacetylases (HDACs), although evidence has emerged that these enzymes also modify the lysine residues of a large variety of non-histone proteins. HDACs have been classified into four groups (I-IV) based on structural homology and function. The class I, II and IV HDACs have a zinc centre to their active site and are inhibited by trichostatin A, these are often referred to as classical HDACs. Class III HDACs lack the zinc atom and are not inhibited by trichostatin A. Class I HDACs have been found in all tissue types and have been most extensively studied in relation to transcription (for a review see Dokmanovic et al. 2007).

1.5.4 Chromatin remodelling

In order to allow the transcriptional machinery access to the DNA it is necessary that the DNA is unbound from its nucleosome. This may occur by the unwrapping of the DNA from the histone octamer or by 'sliding' the nucleosome along the DNA to a new position. Such changes are referred to as chromatin remodelling. Specific cellular chromatin-remodelling complexes exist but chromatin remodelling also occurs in response to transcriptionally active Pol II complexes(Kireeva *et al.*, 2002). It has also been demonstrated that chromatin remodelling complexes can recruit transcription factors to promoters(Utley *et al.*, 1997; Workman and Kingston, 1998).

1.5.5 Initiation of transcription

After transcription factors have bound to the promoter and recruited the remaining subunits of the pre-initiation complex the recruitment of RNA Pol II heralds the start of transcription. RNA Pol II is a DNA dependent RNA polymerase which does not require a primer. After ATP dependent melting of the DNA mediated by TFIIH the enzyme can access the now single stranded DNA of the start site. There the DNA is brought into the active site of RNA Pol II and the initial RNA chain is assembled as a heterodimer with the DNA template(Hahn, 2004). Single molecule studies have suggested that the first 9-11nt of DNA are 'pulled' into the active site by Pol II (Revyakin *et al.*, 2006) and that repeated abortive initiations occur before Pol II escapes the promoter. Escape from the promoter requires phosphorylation of the C terminal domain of Pol II and the movement of the nascent RNA transcript into the RNA exit channel. Transcription has now been initiated but production of mRNA requires the presence of factors that promote

elongation of the RNA. High throughput ChIP studies have demonstrated that transcription from many eukaryotic genes is paused at the point of productive elongation(reviewed in Adelman & Lis 2012) and that this pausing can be reversed by the recruitment of elongation factors. It has been suggested that pausing has emerged as a mechanism to facilitate rapid switching on of transcription(Core and Lis, 2008; Adelman and Lis, 2012).

1.5.6 Elongation and removal of chromatin barriers to transcription

Once Pol II has released the GTFs and moved into the coding region of the gene, recruitment of the cellular elongation machinery is essential to complete transcription. Two negative elongation factors bind to Pol II and prevent elongation, these are DRB sensitivity inducing factor (DSIF) and the negative elongation factor (NELF)(Yamaguchi *et al.*, 1999). The positive transcription elongation factor b (P-TEFb) consists of Cyclin Dependent Kinase 9 (CDK9) and one of cyclin T1, T2 or K. The kinase action of P-TEFb phosphorylates DSIF, NELF and the CTD of Pol II. This phosphorylation removes NELF and DSIF and permits elongation of the RNA transcript(Yamaguchi *et al.*, 1999).

In addition to factors acting on RNA Pol II itself elongation is dependent on factors responsible for modifying the local histone environment. The PAF complex (reviewed in Rosonina & Manley 2005) has been demonstrated by ChIP assays to be present at all actively transcribed genes. PAF is involved in histone K4 methylation, a mark which is associated with active transcription(Krogan *et al.*, 2003; Wood *et al.*, 2003). Other chromatin modelling factors recruited by PAF include FACT and Chd1 which are essential for transcription from chromatinised DNA and are involved in the recruitment of histone acetyl transferases as well as the temporary removal of histone proteins ahead of the advancing Pol II complex(Simic *et al.*, 2003; Winkler and Luger, 2011).

1.5.7 Replacement of histones

As the transcription machinery passes along a gene it appears to be essential for the cell to reset the chromatin architecture onto the transcribed regions. ChIP analysis of histone replacement levels in yeasts after transfer into a glucose

containing growth medium have shown that histone proteins are removed from the actively transcribed genes but deposited again within ten minutes of cessation of transcription(Schwabish and Struhl, 2004). The requirement for the replacement of histones has been linked to a need to prevent transcription from cryptic promoters as mutations in histone chaperones such as FACT result in the generation of transcripts from internal start sites.

1.5.8 Chromatin and HIV

Two nucleosomes, termed nuc-0 and nuc-1, consistently form within the HIV 5' LTR (Verdin, Paras and Van Lint, 1993; Van Lint, Bouchat and Marcello, 2013). Nuc-0 is found at the beginning of the U3 region and nuc-1 250bp downstream of this. The region between the nucleosomes contains binding sites for transcription factors as well as the transcription start site. Modification of the histones in these nucleosomes has been extensively studied in the field of HIV latency.

As with cellular genes histone acetylation has been linked to increases in HIV expression. As discussed above the transactivator protein Tat is acetylated by CRB/p300, these proteins also function as histone acetyltransferases and have been identified at the LTR of transcriptionally active proviruses(Benkirane *et al.*, 1998; Marzio *et al.*, 1998). Specific histone acetylation marks which have been associated with transcriptional upregulation are H3K14, H4K5, H4K8 and H4K16 for nuc-0 and H3K9, H4K8 and H4K16 for nuc-1(Lusic *et al.*, 2003). The ability of inhibitors of histone deacetylases to promote HIV transcription (discussed further below) adds further evidence that acetylation of these histones is essential for transcription of HIV.

In contrast to acetylation, methylation of histone proteins has been linked to repression of the HIV LTR. Tri-methylation of lysine 9 of histone protein H3 is a particularly strong repressive mark(Pearson *et al.*, 2008; Tyagi, Pearson and Karn, 2010) and in other eukaryotic systems is associated with centromeric heterochromatin which is highly compact. In the context of HIV the H3K9 methyltransferase SUV39 has been studied using siRNA knockdown in an LTR-luciferase system(du Chéné *et al.*, 2007) where the knockdown caused increased

transcriptional response to Tat from an integrated LTR. The recent discovery of the Human Silencing Hub (HUSH) has also suggested a role for the H3K9 methyltransferase SetDB1 in HIV latency (discussed in more detail in Chapter 4). In brief; knockdown of HUSH reduced recruitment to SetDB1 and led to increased expression from a number of lentiviral systems as well as reversing H3K9me3 deposition at a number of cellular loci(Tchasovnikarova *et al.*, 2015). This raises the question of whether SetDB1 has a role in HIV latency.

Another methylation mark that has been implicated in silencing of HIV is H3K27me3. The proteins responsible for the deposition of H3K27me3 are the polycomb repressive complex 2 (PRC2). Experiments using siRNA knockdown of one of the principal components of PRC2 (EZH2) (Friedman *et al.*, 2011; Nguyen *et al.*, 2017) in a Jurkat cell based model of HIV latency induced transcription from integrated proviruses as well as sensitising them to exogenous stimulus.

Higher order chromatin structures may have a role in long term silencing of latent HIV although these remain relatively unexplored. One study has explored the requirement for P-TEFb to phosphorylate histone protein H1 which links nucleosomes and aids in compacting chromatin. Using ChIP assays and knockdowns they demonstrated that H1 dissociation from the LTR was dependent on phosphorylation and that this in turn was dependent on P-TEFb(O'Brien *et al.*, 2010).

1.5.9 Chromatin structure does not fully explain HIV latency

The relevance of heterochromatin formation to HIV latency *in vivo* has been challenged by the finding that HIV latency can be established within actively transcribed genes in resting CD4+ cells(Han *et al.*, 2004b). The cellular environment and chromatin modification do not provide a complete explanation for the dynamics of HIV gene expression. Cell line models of HIV gene expression show that populations of non-expressers and high expressers predominate with only a small number of intermediate expressers. If the intermediate expressers are purified, expression decays to give two populations, high and low (Weinberger *et al.*, 2005). This argues for a stochastic element to gene expression where LTR

basal transcription is distributed across a range but must exceed a threshold to establish a positive feedback loop and complete productive transcription. The stochastic element to the control of HIV transcription is confirmed by studies which have shown that compounds which increase 'noise' in cellular expression also reactivate latent HIV(Dar *et al.*, 2014). The modification of 'noise' in cellular systems may identify novel strategies for activating latent viral genomes (Kellogg & Tay 2015)

Methylation of CpG groups in cellular DNA is associated with transcriptional repression(reviewed in Jones 2012). It was initially proposed that DNA methylation would have a role to play in HIV latency however a detailed study using bisulphite sequencing of proviral LTRs in resting CD4+ cells from infected patients receiving antiretroviral treatment did not show an enrichment of DNA methylation marks(Blazkova *et al.*, 2012). This is in contrast to the provirus of HTLV-1 where methylation of the 5' LTR has been shown to regulate transcription(Miyazato *et al.*, 2016).

A further cellular mechanism not involved in chromatin modification but which has been linked to HIV latency is the mTOR complex. This group of related proteins is involved in transcriptional responses in a variety of cellular signalling cascades. Using a genome wide shRNA screen in the J-lat model of HIV latency (discussed below) Eric Verdin's laboratory identified factors which when knocked down reduced the ability of the latent virus to reactivate on stimulation with PMA(Besnard *et al.*, 2016). They showed in both an shRNA and a CRISPR system that knockdown of mTOR prevented exit from latency. They also demonstrated that pharmacological inhibition of mTOR could reproduce the effect. They showed that mTOR was upregulated by CD3/CD28 crosslinking and that its role in HIV latency was to regulate the phosphorylation of CDK9. By reducing the activity of mTOR the rate of phosphorylation of CDK9 was reduced and therefore transcription from the proviral LTR was prevented. They also demonstrated this effect in cells from infected patients.

HIV is known to preferentially integrate into active cellular genes. Previously it was thought that integration into gene deserts would be a cause of HIV latency however sequencing of these integration events has been found them to predominantly reflect integration of defective viruses(Cohn *et al.*, 2015).

1.5.10 Non-induced proviruses

Despite maximal CD4+ T cell activation some proviruses cannot be reactivated. Work done in the laboratory of Robert Siliciano has shown that only a fraction (<1%) of all integration events can be made to produce virus (Ho *et al.*, 2013). They examined integrated proviruses in patient CD4+ T cells, which did not produce detectable p24 on stimulation with PHA. Mutations were identified in 88.3% of non-induced proviruses, which would render the virus defective. The remaining 11.7% contained full-length proviruses, which appeared replication competent. By reconstructing a number of full-length sequences these were shown to be replication competent. This suggests that for each virus reactivated by PHA stimulation there are approximately 60 replication competent viruses, which are not induced. This implies that any attempt to eliminate the latent viral reservoir by the 'shock and kill' method is likely to require a more efficient activating strategy.

1.6 Model systems for studying HIV latency

The rarity of latently infected CD4+ T cells *in vivo* has necessitated the development of model systems in order to study the mechanisms involved in silencing as well as to study strategies for reactivating silenced proviruses. Broadly speaking the field can be divided into models based on cell lines and those based on primary cells.

1.6.1 Cell line models of HIV latency

One of the first cell line models to give insights into the phenomenon of HIV latency was the ACH-2 cell line. In one of the first descriptions of the phenomenon of HIV latency A2.01 (a variant of the CEM T cell line) cells were infected with the replication competent laboratory virus HIV_{LAV} in limiting dilution and then

subcloned after allowing time for the virus to kill infected cells(Folks *et al.*, 1986). The lab identified a group of cells, which had low levels of surface CD4 suggesting they had been infected but had survived the infection. Folks' group demonstrated that replicating virus could be released from these cells by addition of iododeoxyuridine to the culture medium (Folks *et al.*, 1986) which had been used elsewhere to induced endogenous retroviruses and later showed the virus release could be induced by Tumour Necrosis Factor(Folks *et al.*, 1989). Further examination of the provirus in the ACH-2 clone demonstrated a point mutation in the TAR sequence that reduced the efficiency of Tat binding(Emiliani *et al.*, 1996) and suggested that the cell line may not be a faithful representation of the *in vivo* situation. More recently a study of integration sites in model systems of HIV latency has demonstrated that while the majority of ACH2 cells have a single provirus there is a significant minority with alternate integration sites suggesting low level replication occurs without induction(Symons *et al.*, 2017). These findings together mean that the results of studies based entirely on ACH-2 cells are limited in their generalisability.

Another cell line model which has been widely studied is the J-lat model(Jordan, Bisgrove and Verdin, 2003). This model was developed in Eric Verdin's laboratory and was based on the Jurkat T cell line. It utilised fluorescent vectors for the first time in a latency model. The Verdin group infected Jurkat cells with vector constructs encoding LTR-*tat*-IRES-GFP. Expression of GFP was determined by flow cytometry; cells, which were GFP positive, were deemed to represent productive infection and were removed by cell sorting. The negative population was treated with TNF and a small fraction expressed GFP. These cells were sub-cloned and shown to reproducibly become GFP positive on stimulation. They also demonstrated undetectable HIV RNA without stimulation and showed that each clone harboured a single HIV provirus.

In the same paper the Verdin group described the production of a similar set of clones using a more extensive HIV based vector which encoded a near full-length virus rendered non infectious by a deletion in the envelope region and with Nef substituted by GFP. Clones made using this provirus behaved similarly to those

with the truncated construct but also produced detectable HIV proteins upon stimulation meaning that induction of transcription could be monitored by p24 level in the supernatant in addition to the detection of GFP by flow cytometry. J-lat cells and similar Jurkat and vector based systems have since been used to study the response of proviruses to putative latency reversing agents as well as to deliver insights into the biology of HIV latency(Bisgrove *et al.*, 2007; Friedman *et al.*, 2011; Nguyen *et al.*, 2017).

1.6.2 Primary cell models of HIV latency

Despite the advantages of cell line models, i.e. that they can be cultured indefinitely and that sub-clones bearing a single integration site can be selected out, there are concerns about how physiologically relevant they can be. Therefore there has also been a move to generate model systems for studying HIV silencing in primary cells. Direct infection of resting CD4 cells *in vitro* is difficult and yields low rates of productive infection(Stevenson *et al.*, 1990). Nonetheless model systems have been developed which rely on spinoculation (infecting cells while in a centrifuge at 800-1200 x *g*) of resting CD4+ cells (by their definition cells which are CD69 and HLA DR negative). One of the first (Swiggard *et al.*, 2005) utilised the laboratory virus IIIB and demonstrated that cells infected this way would permit integration of proviral DNA without altering the activation state of the cells. They also showed that on stimulation of the T cell receptor by CD3/CD28 crosslinking intracellular HIV p24 could be detected by flow cytometry.

A modification of this system published by the Greene group (Lassen *et al.*, 2012) was to use $HIV_{NL4.3}$ derived viruses which, in place of Nef, expressed a GFP-IRES-Nef cassette, luciferase alone or mCherry-T2A-luciferase. In this system the fluorophores GFP and mCherry allow quantitation of the number of cells transcribing from the LTR using flow cytometry. The luciferase allows determination of the strength of the LTR driven signal; luciferase cleaves a variety of substrates and produces light allowing quantification of enzyme activity from lysed cells by light output which can then be correlated with the amount of LTR driven transcription.

The Planelles model (Bosque and Planelles, 2009, 2011) also uses fluorescent viruses. In this model naïve (CD4+,CD45RA+, CD45RO−,CCR7+,CD62L+,CD27+) T cells are activated in the presence of cytokines (TGF- β, IL-4, IL-12) chosen to drive proliferation of the cells without Th1 or Th2 polarisation before being infected at day 7. The authors equate the phenotype of these cells with central memory T cells. Once the cells are activated again at day 14 with CD3/CD28 crosslinking they observe an increase in cells expressing viral genes however they do not observe a strong response to HDAC inhibitors or PKC agonists.

Siliciano's model (Yang *et al.*, 2009) takes CD4+ T cells and transduces them with a vector expressing the anti-apoptotic protein Bcl2 and then cultures them in the absence of cytokines for 3-4 weeks to enable transition back to a resting phenotype. The cells are stimulated again and infected with an HIV derived virus expressing GFP in place of *env* before being returned to culture for a further 4 weeks. At this point GFP negative cells are sorted and stimulated with the chosen activating agent and the proportion of GFP positive cells is taken as the read out of reactivation from latency.

The Lewin model utilises a system of detection similar to the viral outgrowth assays used in clinical studies of HIV latency (discussed below). HLA DR-, CD 69-, CD25-, CD4+ T cells are isolated and cultured with the chemokine CCL19 before being infected with $HIV_{NL4.3}$. Treating the cells with CL19 is believed to stabilise the integrase complex and therefore increase the rate of completed integration in the resting cells(Saleh *et al.*, 2007, 2016). Four days after infection the cells are seeded into 96 well plates with CD8 depleted PBMC as feeder cells. The stimulating agent is added to the supernatant and released virus multiplies in the PBMC where it can be detected as soluble reverse transcriptase activity

The model developed by Spina and colleagues (Spina, Guatelli and Richman, 1995) also utilises CD4+ T cells. The whole CD4+ T cell fraction is divided into two with one group of cell maintained in culture without stimulation and the other stimulated with CD3/CdD28 crosslinking and infected with $HIV_{NL4.3}$. The infected

dividing cells are then co-cultured with the unstimulated cells for 3 days. After co-culture the non-dividing cells are isolated by flow cytometry. The cells are then reactivated in the presence of the integrase inhibitor raltegravir and reactivation measured as copies of HIV RNA per cell.

1.6.3 Comparison of model systems with primary cells

To date only one study has compared the performance of the various models described above to patient CD4+ cells activated *ex vivo*. This study (Spina *et al.*, 2013) examined the response of each model to a variety of latency reversing stimuli. The outputs were normalised to the maximum signal observed enabling comparison between models, patient cells were also treated with a more limited range of stimuli. In an analysis by clustering the Lewin model and J-lat 5A8 cells were found to display a reactivation pattern most similar to outgrowth from patient cells. The Spina model on the other hand was found to be the least similar to patient cells.

1.7 Latency reversing agents and a 'shock and kill' approach

Reversing the latency phenotype has become a central goal of strategies to achieve HIV cure. The principle rests on the idea that since latent viruses are not vulnerable to the immune system or to antiretroviral therapy; by rousing the virus from latency it may destroy its host cell cytopathically or render it susceptible to immune clearance either of which would lead to the latent reservoir being depleted. The strategy of activating the latent virus and boosting the immune response to it has been termed 'shock and kill'(Deeks, 2012). This approach has led to the development of a number of pharmacological approaches to disrupt HIV latency.

Acetylation of the histone proteins within nuc-1 is required to allow productive transcription to occur(Lusic *et al.*, 2003). Inhibitors of histone deacetylases (HDACi) have been shown to be able to induce virus outgrowth from latently infected patient cells(Archin, Espeseth, *et al.*, 2009). This effect has been found to be restricted to inhibitors of class I histone deacetylases(Archin, Keedy, *et al.*, 2009). In the same study it was found that these agents did not cause cellular

activation. A number of histone deacetylase inhibitors are already licenced for use in humans; valproic acid for epilepsy, vorinostat (SAHA) and romidepsin for cutaneous T cell lymphomas and panobinostat for multiple myeloma. Of these panobinostat has been shown to be the most potent HDAC inhibitor *ex vivo*(Rasmussen *et al.*, 2013).

Another therapeutic target for potential latency reversing agents is increasing the availability of PTEF-b to promote completion of transcription from the HIV LTR. One such compound is hexamethylene bisacetamide (HMBA) which releases P-TEFb from the 7SK snRNP where it is bound to HEXIM1(Barboric *et al.*, 2007; Contreras *et al.*, 2007; Krueger *et al.*, 2010). HMBA has been shown to disrupt HIV latency in CD4+ T cells by allowing initiated transcripts to progress to completion(Choudhary, Archin and Margolis, 2008). PTEF-b also interacts with bromodomain containing protein BRD4 which prevents efficient transactivation by Tat(Bisgrove *et al.*, 2007). The bromodomain inhibitor JQ1 can disrupt this interaction and has also been shown to reactivate virus from latency(Banerjee *et al.*, 2012; Boehm *et al.*, 2013).

The NFκB family of proteins promote transcriptional elongation of cellular genes, and also act on HIV transcription to promote productive elongation. In the absence of activating signals NFκB is sequestered in the cytoplasm as inactive homodimers. To become active it must form heterodimers made of class I and class II NFκB proteins. This process is enhanced by the protein kinase C family of enzymes. The protein kinase C/NFκB pathway's role in upregulating transcription suggests a potential to disrupt HIV latency by increasing the availability of NFκB. PKC agonists (prostratin, bryostatin) have therefore been examined as anti-latency compounds and shown to increase HIV gene expression both in cell models and *ex vivo* primary cells(Mehla *et al.*, 2010; Díaz *et al.*, 2015).
The response of integrated HIV to a range of cellular transcription factors means that model systems can exploit the large burst of transcription factors that are released on activation of T cells. The leading strategy for activating T cells is crosslinking of the T cell receptor. This is achieved either with antibody-coated beads targeting CD3 and CD28 or with the plant derived lectin

phytohaemagglutinin. While these strategies cause efficient reactivation of latent HIV *in vitro* their clinical use is severely limited. The global activation of T cells would cause unacceptable hypersensitivity reactions.

Other approaches to reversing HIV latency have been identified by the screening of libraries of known drugs against primary cell models of HIV latency. The advantage being that existing drugs can be more readily trialled in humans as they will already have a history of use. One such 'hit' has been the identification of disulfiram as a potential latency reversing agent. It was shown to increase expression of GFP in an HIV LTR driven system in Bcl-2 transduced primary cells(Xing *et al.*, 2011). Although the mechanism by which this occurs is not known disulfiram is licensed for the treatment of alcohol abuse and has already been translated into clinical trials of latency reversal.

In isolation latency reversing agents should cause an increase in virus production however it is not clear that the cytopathic effect of reactivating a virus from latency will be sufficient to cause a depletion of the latent viral reservoir. To augment the depletion of the reservoir it is proposed that a strong immune response to viruses that are activated from latency is required. Studies examining cytotoxic lymphocyte (CTL) responses to reactivated latent viruses have found that escape mutations making CD8+ T cell responses less effective are archived early in the course of infection(Deng *et al.*, 2015). Nonetheless it has been possible to stimulate CTL responses *ex vivo* to reactivated latent viruses (Shan *et al.*, 2012; Deng *et al.*, 2015). Strategies to induce optimum CTL responses after treatment with a latency reversing agent are currently under investigation but include vaccination given at the time of latency reversal to boost the CTL response to the proteins produced.

Alternative approaches to the reservoir

An interesting recent proposal has been to enhance silencing of latent proviruses in order to preclude reactivation and therefore prevent rebound viraemia after stopping treatment. This approach has been termed 'block-and-lock' by its authors(Kessing *et al.*, 2017). This group have studied the use of a specific

inhibitor of HIV Tat called didehydro-cortistatin A (dCA) in a humanised mouse model of HIV latency as well as in patient CD4+ T cells. They demonstrated reduced virus production from dCA treated cells as well as an increased time to rebound viraemia in mice taken off antiretroviral therapy. It is as yet unknown whether an approach like this could achieve the permanent silencing which would be required for lasting drug-free remission in human subjects.

There has also been significant interest in the use of the gene editing tool CRISPR/Cas9 to remove HIV genomes from infected cells. In brief, the system uses a lentiviral vector to transduce a cell with genes expressing the Cas9 enzyme along with specific guide RNAs. The guide RNA directs the Cas9 to a specific DNA sequence where it cleaves the DNA(for a review see Sander & Joung 2014). By designing guide RNAs specific for HIV sequences it has been possible to excise HIV proviruses from infected cells(Kaminski *et al.*, 2016). The scalability of this approach and the low probability of successfully transducing all infected cells may limit this approach however one effect of transducing cells with an anti-HIV CRISPR system is that the cells then become resistant to infection as any new HIV integrations are targeted by Cas9(Liao *et al.*, 2015). This opens up the possibility of transducing patient cells *ex vivo* and then reintroducing them leaving the patient with a population of HIV resistant T cells. CRISPR/Cas9 mediated activation of latent proviruses has been reported by using LTR specific guide RNA to deliver a Cas9 lacking its endonuclease activity (dCas9) fused to a transactivator or acetyltransferase. The authors demonstrated that this system could induce transcription in a reporter system and the system enhanced the effect of the histone deacetylase inhibitor vorinostat(Limsirichai, Gaj and Schaffer, 2016).

1.8 The *in vivo* reservoir and eradication studies

1.8.1 Measures of the latent reservoir *in vivo*

Viral outgrowth assays

The number of resting CD4+ T cells harbouring replication competent viruses has been considered the most accurate measurement of the latent pool as it is thought that these cells are the source of the virus which rebounds after cessation of antiviral therapy (Brennan *et al.*, 2009; Anderson *et al.*, 2011). The demonstration that a population of CD4+ HLA-DR- T cells from patients on effective treatment could be made to release virus upon stimulation with phytohaemagglutinin (PHA) led to efforts to quantify the latent reservoir *ex vivo* (Finzi *et al.*, 1997; T.-W. Chun *et al.*, 1997; Wong *et al.*, 1997). These studies established the principle of quantitation by seeding resting CD4+ T cells in limiting dilution before activating the cells with a combination of PHA and irradiated allogeneic peripheral blood mononuclear cells (PBMCs). Due to the low frequency of latently infected cells it was necessary to obtain a large volume blood sample (approximately 200 ml) from study participants.

The amount of HIV Gag protein released by reactivating a single latently infected cell is below the limit of detection of a standard p24 ELISA. Therefore the released virus is amplified by co-culture with permissive cells. In the original assay the cells were co-cultured with PHA stimulated, CD8-depleted PBMCs from an HIV-negative donor. Maximum likelihood statistics became the standard to estimate the frequency of latently infected resting CD4+ T cells, expressed as infectious units per million cells (IUPM) (Myers, McQuay and Hollinger, 1994).

The use of allogeneic CD8-depleted PBMCs to amplify outgrowth virus has biological and practical limitations. PBMCs from different donors are not equally permissive for HIV infection, leading to different IUPMs depending on the PBMC donor. This problem is particularly acute when amplifying R5 tropic virus as expression levels of CCR5 vary widely between donors. On the other hand CXCR4 tends to be highly expressed, meaning detection of X4 tropic virus is less problematic.

One solution is pre-screening of donor PBMCs for CCR5 expression level and selecting for high CCR5-expressing donors only. Another option is using cells from matched donor-patient pairs that are shown *in vitro* to exhibit robust viral replication, a method that also requires extensive pre-screening prior to the actual

assay (Archin *et al.*, 2008). Although donor matching reduces inter assay variability for samples obtained from the same HIV-positive donor, it makes the procedure more labour intensive; and together with the need to perform leukapheresis to obtain sufficient latently infected cells, makes it harder for VOA to scale up to meet the needs of a clinical trial.

A number of modifications have been introduced to improve the original viral outgrowth assay. Two new methods utilise cell lines engineered to express high levels of CCR5 (Laird *et al.*, 2013; Fun *et al.*, 2017). These cell line based assays have been shown to give IUPM readings similar to PBMC based assays (Laird *et al.*, 2013; Fun *et al.*, 2017) and to give highly consistent results when sampling the same patient repeatedly. Cell line based systems are also less labour intensive as the cells do not have to be freshly prepared from blood and stimulated with PHA for each assay. One assay uses a custom antibody cocktail to purify resting CD4+ T cells from PBMCs of HIV positive donors in a single step, significantly reducing processing time from approximately, a huge advantage when scaling up assay numbers for clinical trials. Another recent modification proposed to reduce the work required for each assay is stimulating the resting CD4+ T cells with CD3/CD28 microbeads rather than PHA and irradiated allogeneic PBMCs. Although the results correlated well with those derived from the standard stimulus, the microbead based method appeared to be less sensitive (Kuzmichev *et al.*, 2017).

Recently a number of strategies have been proposed for optimising the outgrowth assay by improving the sensitivity of the detection system or simplifying the read-out. Alternatively, use of a more sensitive PCR based detection method instead of the p24 ELISA can reduce the co-culture time required to detect replication-competent HIV (Laird *et al.*, 2013). The TILDA assay (Procopio *et al.*, 2015) utilises the detection of an increase in multiply spliced HIV RNA in CD4+ T cells after stimulation. Similar to the VOA, the cells are seeded in limiting dilution to obtain a quantitative measure. It uses total rather than resting CD4+ T cells, which has the advantage of requiring a smaller blood volume than a VOA and it does not need an extended period of culture for detection, however this means that it measures the inducible reservoir in a population of cells different to that used in the VOA.

Another new assay, TZA, utilizes TZM-bl cells (Sanyal *et al.*, 2017). These cells produce β-galactosidase in response to viral Tat, enabling the presence of HIV to be determined by the enzymatic cleavage of a luminescent compound. In contrast to TILDA the TZA assay utilises resting CD4+ T cells but also requires a significantly smaller blood volume than the standard VOA. Neither of these assays correlated closely with IUPM measured by a standard VOA implying that they may not measure the same component of the HIV reservoir. Another caveat is that as neither assay requires viral replication to detect reactivated virus neither provides unequivocal proof that the detected viral products are derived from replication-competent virus.

Another novel technique that significantly shortens the assay time of the VOA, by use of an enhanced sensitivity of p24 detection at ultra-low concentrations, has recently been applied to HIV latency research (Passaes *et al.*, 2017). This ultrasensitive assay utilises Simoa single molecule automated array technology (Rissin *et al.*, 2010) which can measure sub-femtomolar concentrations of analytes using a micro-bead based antibody capture system. This promising method allows the detection and quantitation of p24 produced by a single reactivated infected CD4+ T cell (Passaes *et al.*, 2017). Although this method is very powerful in detecting viral protein production very early after reactivation, it is not yet clear if the measure represents replication-competent virus and how it correlates to the VOA. Therefore, the results of ongoing clinical studies like RIVER which employ both a VOA and the ultrasensitive p24 assay, among many other measures of the latent HIV reservoir will give useful data on the comparison of these measures.

As efforts to cure HIV progress it may become necessary to be able to detect ever smaller reservoirs with greater confidence. A recent variation on the VOA has been to use humanised mice as the outgrowth system (Charlins *et al.*, 2017). This system was able to give a semi-quantitative estimate of the reservoir size and detected outgrowth from patient cells where a standard VOA had been negative. Although this method is clearly not easily scalable it offers a means of monitoring the reservoir size in those with a negative standard VOA.

Proviral DNA

An alternate and less labour intensive approach to measuring the latent reservoir is to measure HIV DNA levels in circulating CD4+ T cells. These methods utilise quantitative PCR and methods have been described using either normalisation to cellular genes or digital droplet PCR. Efforts to quantify DNA proviral load began shortly after the introduction of HAART made it possible to suppress plasma HIV RNA levels below the limit of detection of a standard assay. In these early studies it was observed that over a year a modest decay in the proviral load occurs (Debiaggi *et al.*, 2000; Ngo-Giang-Huong *et al.*, 2001).

Longer term studies have also found a slow decline in HIV proviral DNA levels(Besson *et al.*, 2014; Luzuriaga *et al.*, 2014) in patients on treatment. It has been demonstrated that lower HIV proviral DNA loads are associated with longer times to viral rebound and to lower set point viral loads after treatment interruption(Yerly *et al.*, 2004; Komninakis *et al.*, 2012). This association with rebound viral load has also been seen in a therapeutic vaccine trial(Li *et al.*, 2014) where high DNA loads were correlated with higher rebound viral set points.

The obvious advantage of a DNA based approach is the ease with which it can be scaled up to process large numbers of samples. Furthermore it has the advantage of requiring a much smaller blood volume than that required for viral outgrowth assays.

One caveat of a molecular approach to measuring the reservoir is that these methods do not distinguish between replication competent and defective proviruses. Since the rebound observed on cessation of antiretroviral therapy must arise from a replication competent provirus this is an important distinction to make. It is interesting to note that HIV proviral DNA has been detected in tissue samples of the so called 'Berlin patient' who has remained aviraemic following an allogeneic stem cell transplant from a CCR5 deficient donor. This is in contrast to another case report of a child who maintained viral suppression following interruption of treatment who had no detectable HIV DNA(Persaud *et al.*, 2013).

1.8.2 Eradication studies in humans

As discussed above, long-term therapy with antiretroviral drugs does not lead to eradication of HIV even where there has been a long history of viral suppression. However studies of the timing of the initiation of therapy have lent hope that a 'functional cure' is possible where a person would remain infected however would not need to take medication. A key study has been the ANRS VISCONTI study(Sáez-Cirión *et al.*, 2013). This study identified a group of patients who had started antiretroviral therapy early in the course of their infection and who on stopping therapy experienced a period of sustained virological suppression. They identified 14 patients who had a median of 89 months of undetectable viraemia after stopping treatment. Six patients became viraemic over the course of the study. Interestingly the HLA*B27 and HLA*B57 alleles which are associated with spontaneous control were not found to be overrepresented in the post treatment controllers. They also found lower and declining levels of cell associated HIV DNA in post treatment controllers.

Similarly the SPARTAC trial (Fidler *et al.*, 2013) identified a correlation between the length of time to viral rebound after cessation of treatment and the timing of initiation of treatment. They found that those who started treatment earlier after primary infection had a longer time to rebound. There has also been a number of high profile cases in which individuals started on therapy early had a sustained remission on stopping treatment (Persaud *et al.*, 2013). These studies suggest that early and aggressive anti-HIV treatment may lead to a smaller reservoir of virus that in turn slows the rate of viral rebound.

In addition to studies of the timing of administration of antiretroviral therapy pilot studies examining the use of candidate latency reversing agents have been carried out. One of the first drugs to be trialled, vorinostat is an HDAC inhibitor, working by inhibiting the deacetylation of histones and promoting transcription of latent HIV genomes. Vorinostat has been found to increase cell associated HIV RNA in patients(Archin *et al.*, 2012) however these effects were not found to translate into decreases in the measured HIV reservoir(Archin *et al.*, 2014).

Similarly the HDAC inhibitor panobinostat was found to cause a transient increase in viraemia but no change in measures of the reservoir(Rasmussen *et al.*, 2014).

A trial of the putative anti-latency compound disulfiram (Elliott *et al.*, 2015) demonstrated its safety in a small group of patients and found that cell associated RNA increased by up to 2.5 fold during administration of disulfiram indicating that HIV transcription (presumed to be from latent genomes) was increased

Despite these early studies it remains unclear exactly which is the optimum method for reactivating latent viruses *in vivo*. Furthermore what measure of the reservoir correlates best with clinically useful outcomes remains unknown. Larger studies with longer follow up periods will be essential to answering these questions. It will also be necessary for studies to include periods of treatment interruption to evaluate the success of any intervention in preventing relapse. This raises an ethical issue as study participants will be exposing themselves to the risks of worsening immune function and emerging viral resistance by stopping their antiretroviral drugs. At present there are insufficient trial data to quantify these risks.

Chapter 2: Materials and methods

2.1 Plasmids

The following plasmids were used in this project:
pNL4.3.deltaENV.GFP (Aidsreagents, no. 11100)
pNL4.3 (Aidsreagents, no. 114)
pRGH (Aidsreagents, no. 12427)
pCMV-VSVG (Addgene)
pCMV-deltaR8.91 (Addgene)
pBR322 (Promega)

The shRNA expressing vector plasmids described in chapter 4 were a kind gift of the Lehner laboratory

2.1.1 Transformation of competent cells

Transformations were carried out using XL10 Gold competent cells (Agilent). 1µl of β-mercaptoethanol was added to 45µl of competent cells in a 14ml round bottom 'falcon' tube (Corning) and incubated on ice for 10 minutes. 0.1-50ng of plasmid was added and incubated for 30 minutes on ice. The mixture was then heat shocked at 42°C for 30 seconds in a water bath then returned to ice for 2 minutes. 450µl of pre-warmed NZY$^+$ broth (10g/l NZ amine, 5g/l yeast extract, 5g/l NaCl, 12.5mM $MgCl_2$, 12.5mM $MgSO_4$, 20mM glucose, pH 7.5, autoclaved) was added and the cells were incubated for 1 hour at 37°C in a shaker. Cells were then plated on an LB agar (10g/l tryptone, 10g/l NaCl, 5g/l yeast extract, 15g/l agar, autoclaved) with 100µg/ml ampicillin (Sigma) and incubated at 37°C. Plates were examined for colonies after 18 hours.

2.1.2 Small scale extraction of plasmids from cultured bacteria

Colonies from transformed cells were picked and transferred to 3ml aliquots of lysogeny broth (LB, 10g/l tryptone, 10g/l NaCl, 5g/l yeast extract, autoclaved) containing 100µg/ml ampicillin and cultured overnight in a shaker incubator at 37°C. 1ml of culture was taken for small-scale plasmid extraction, the remaining culture was reserved to be used to start larger cultures if required. Small-scale plasmid extraction was carried out using a Qiagen plasmid mini kit according to the manufacturer's instructions. The kit works by the alkaline lysis of the bacteria,

the precipitation and removal of proteins, the precipitation and binding of DNA to a silica membrane by high salt concentrations and the elution of the DNA from the membrane. The process is described below.

Bacteria were pelleted by centrifugation at 6000x*g* for 5 minutes in a bench top microcentrifuge and the culture medium discarded. Pelleted bacteria were resuspended with 250µl of buffer P1 (50 mM Tris pH8.0, 10 mM ethylenediaminetetraacetate (EDTA), 100 mg/ml RNaseA) then lysed with 250µl of buffer P2 (200mM NaOH, 1% sodium dodecyl sulphate (SDS)). Lysis was stopped by 300µl of buffer P3 (3M potassium acetate pH 5.5). The sample was centrifuged at 16000x*g* on a desktop centrifuge for five minutes to pellet the precipitated protein and genomic DNA. The supernatant was transferred to a column and centrifuged for 1 minute allowing the DNA to bind to the column. The flow through was discarded. 700µl of buffer PE (10mM tris(hydroxymethyl)aminomethane (Tris) -HCl pH 7.5, 80% ethanol) was added and the column centrifuged to wash the DNA. After discarding the flow through the column was returned to the centrifuge for a further minute to remove any residual alcohol and dry the membrane. Finally the DNA was eluted by placing the column into a clean microcentrifuge tube, adding 50µl water to the membrane and centrifuging for 1 minute. DNA was either analysed immediately or stored at -20°C.

2.1.3 Larger scale preparation of plasmid

To prepare larger quantities of DNA of high purity 500ml of LB both containing ampicillin was inoculated with 1ml of starter culture prepared as above. The culture was incubated overnight at 37°C in a shaker incubator. The bacteria were pelleted and lysed and the DNA was extracted using the Qiagen plasmid maxi kit which functions on the same principles as above but with larger volumes of buffer and a higher column capacity.

2.1.4 Agarose gel analysis

PCR products, plasmids and digests were analysed using agarose gel electrophoresis. Gels were made by the addition of 0.8-2% w/v agarose to 1x tris-

borate-ethylenediaminetetraacetate (TBE, 5x solution: 54g/l tris base, 27.5g/l boric acid, 20ml of 0.5M EDTA, pH8.3) and stained with 0.1µl/ml of the fluorescent DNA intercalating stain SYBRsafe (Thermo Fisher). Higher percentage gels were used for resolution of smaller fragments. Gels were cast in an appropriately sized tank with a comb to create wells for loading DNA. DNA was mixed with a loading buffer (Gelpilot, Qiagen) to increase the sample density and contains a dye to visually estimate the progress of the DNA through the gel. 1x TBE was also used as the running buffer for the electrophoresis tank. Electrophoresis was carried out at 75V for 30-60 minutes depending on the size of fragment to be visualised. Identification of band sizes was made by loading an aliquot of a ladder of DNA fragments of known sizes (Hyperladder I and Hyperladder IV, Bioline).

2.1.5 DNA fragment extraction from agarose gels

Following electrophoresis DNA in the agarose gel was visualised on an ultraviolet light box. Bands to be extracted were excised from the gel using a scalpel, taking as little agarose as possible. The DNA was extracted using the Qiagen gel extraction kit. The extracted gel piece was weighed to allow calculation of buffer volumes 100mg = 100µl. 3 volumes of buffer QG (5.5M guanidine thiocyanate, 20mM Tris-HCl pH6.5) were added to the gel slice and incubated at 50°C for 10 minutes until the gel had dissolved. 1 volume of isopropanol was added to the sample. The solution was then transferred to a silica column similar to those used for plasmid DNA extraction above. The sample was then centrifuged at 16000x*g* for 1 minute to allow the solution to pass through the column and the DNA to bind. 700µl of buffer PE was added and the column centrifuged to wash the DNA. The column was returned to the centrifuge for a further minute to remove any residual alcohol and dry the membrane. DNA was then eluted by adding 50µl and centrifuging the column for 1 minute in a clean microcentrifuge tube.

2.1.6 Quantitation of DNA

DNA concentration was determined by spectroscopy. Water without DNA was used as the blank in order to calibrate the instrument. The absorbance of the sample was determined by the spectrometer (Nanodrop, Thermofisher). The

concentration of DNA was determined by the light absorbance at 260 nm. The purity of DNA was determined by the ratio of light absorbance at 260 nm over that at 280 nm.

2.1.7 Restriction digest of DNA

Digestion of DNA by bacterial restriction enzymes was used for the purposes of cloning and for confirming the identity of a plasmid by the creation of a specific pattern of fragments. Bacterial restriction enzymes cleave DNA at sequence specific loci and may leave either blunt ends or 'sticky' ends where one strand of DNA is longer creating an overlap that readily binds a complementary sticky end. Digestion reactions were made up as follows:

DNA 3-5µl
Restriction enzymes 0.5µl (each)
10x Multicore buffer 1µl (Final buffer concentration - 25mM Tris-acetate pH 7.5, 100mM potassium acetate, 10mM magnesium acetate, 1mM dithiothreitol (DTT))
H_2O to make 10µl

Reactions were incubated at 37°C for 90 minutes.

2.1.8 Ligation of DNA fragments

DNA fragments were enzymatically joined using T4 ligase (Promega), an enzyme that creates phosphodiester bonds between the 3' hydroxyl end of one DNA strand and the 5' phosphate group of the other. The reaction consumes two molecules of ATP per bond formed. To perform a ligation reaction the concentration of DNA in the backbone and insert samples was determined as above. Samples were then mixed in a ratio of 1 backbone molecule to 3 insert molecules. Ligation reactions were made up as follows:

Backbone DNA
Insert DNA
10x ligase buffer 1µl (500mM Tris-HCl pH 6.8, 100mM $MgCl_2$, 100mM DTT, 10mM ATP)
Ligase 0.5µl
H_2O to make 10µl

Reactions were incubated overnight at room temperature. Reaction mixes were transformed into bacteria as described above.

2.2 RNA extraction and reverse transcription

2.2.1 RNA extraction from cultured cells

RNA was extracted from cells using the Qiagen RNeasy mini plus kit. This kit utilises an additional column step to remove genomic DNA from the preparation. This is in lieu of performing a DNase digest which can reduce yields due to the extra handling of the RNA. An alternative to commercial kits is phenol chloroform extraction although this was not utilised in this project.

$1x10^6$ cells were pelleted and then resuspended in 350µl buffer RLT (the exact composition of the buffers in this kit are not specified by the manufacturer). The suspension was then transferred to the proprietary gDNA eliminator column and centrifuged at 16000x*g* for 30 seconds to bind the DNA to column. The column was discarded and 350µl of 70% ethanol added to the flow through. This was then transferred to an RNA binding column and centrifuged for 30 seconds. The RNA bound to the column was washed with 700µl buffer RW1 then twice with 500µl of RPE, centrifuging for 15 seconds after the addition of each buffer. Finally the RNA was eluted with 50µl of RNase free water by centrifugation into a clean microcentrifuge tube. Samples which did not proceed directly to reverse transcription were immediately frozen at -70°C.

2.2.2 Reverse transcription of RNA

Complementary strand DNA (cDNA) was made from RNA using the High Capacity cDNA kit (Thermofisher). The kit contains a reverse transcriptase, RNase inhibitor, random primers and a buffer concentrate. The random primers permit cDNA production from any input RNA. It is possible to perform the reaction with specific primers if a particular cDNA product is sought. A 2x mastermix was made up as below and scaled according to the number of reactions. To reduce the effect of pipetting errors an extra reaction's worth of reagents was included each time.

2x mastermix for a single reaction:
10x RT buffer 2µl
100mM dNTPs 0.8µl

10x Random primers 2µl
Reverse transcriptase 1µl
RNase inhibitor 1µl
RNase free H_2O 3.2µl

To 10µl of this mix was added 10µl of RNA as extracted above. For each experiment an identical reaction was set up using 4.2µl water and no reverse transcriptase. Reaction mixes were incubated in a thermal cycler for two hours at 37°C followed by 5 minutes at 85°C. After conversion to cDNA samples were stored at -20°C or proceeded to clean up prior to PCR.

2.2.3 Clean-up of RT products

After reverse transcription cDNA products were cleaned up using the Qiagen PCR clean up kit. This kit uses similar technology and protocols to the gel extraction kit described above to remove dNTPs and enzymes from the mix and to bind the cDNA to a column from which it can be eluted.

2.3 Polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify cDNA formed from HIV mRNAs extracted from HIV infected cells. Primers were designed using the online software Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast). For the purposes of this project one primer set was used for the detection of HIV transcripts by PCR and a second set for the detection of beta actin for normalisation of the signal. The primers (Sigma) were as follows

HIV 640F – 5'-GCCCGAACAGGGACTTGAAA-3'
HIV 5885R – 5' TTGGCTGACTTCCTGGATGC-3'

Beta actin forward – 5'-CCTCGCCTTTGCCGATCC-3'
Beta actin reverse – 5'-GGATCTTCATGAGGTAGTCAGTC-3'

For each experiment a PCR mastermix containing the primers, polymerase, buffer and dNTPs was made up containing sufficient material for all the reactions. For each reaction:

Forward primer 0.25µl
Reverse primer 0.25µl
GoTaq Polymerase (Promega) 0.25µl

dNTPs (Promega) 1µl
Buffer (Promega) 10µl
Template 5-10µl
H_2O to make 50µl

Reactions were thermal cycler for programme	x35	92°C	2 minutes	loaded into a the following
		92°C	1 minute	
		58°C	30 seconds	
		72°C	1 minute	
		72°C	5 minutes	
		4°C	Hold	

2.4 Quantitative PCR

Quantitative PCR utilises the detection of fluorescence from a probe to measure the abundance of a DNA molecule in the mix. The probes have a fluorescent marker at one end and a quencher at the other. The fluorescence from unbound probes is quenched as the single stranded probe keeps the quencher close to the fluorophore. When DNA polymerase encounters a bound probe the probe is degraded and the fluorophore is released into solution where its fluorescence can be detected. Quantitative PCR was carried out with TaqMan qPCR Mastermix (Applied Biosystems); containing the polymerase enzyme and buffer; following manufacturer's instructions. The reactions were analysed on StepOnePlus real time PCR system (Applied Biosystems) using the following cycling parameters: 1 cycle at 95°C for 8 minutes, 45 cycles at 95°C for 10 seconds followed by 60°C for 1 minute. The primers and probes (Sigma) used were:

Tat1 forward - 5'- AGA TCT CTC GAC GCA GGA CT -3'
Tat1 reverse - 5'- GGC TGA CTT CCT GGA TGC TT -3'
D1A3 probe (tat1) - 5' - [6FAM] - TCG ACA CCC AAT TCA GTC GC - [TAM] - 3'

Tat2 forward - 5' - GGA CAG CAG AGA TCC AGT TTG - 3'
Tat2 reverse - 5' - GAT GCT TCC AGG GCT CTA GTC - 3'
D2A3 probe (tat2) - 5' - [6FAM] - GTC GAC ACC CAA TTC TTT CCA G - [TAM] - 3'

All-tat/vpr forward - 5' - TCC TAT GGC AGG AAG AAG CG - 3'
All-tat/vpr reverse - 5' - AGC TTG ATG AGT CTG ACT GT - 3'

All-tat/vpr probe - 5’ [6FAM] TCT GAT GAG CTC TTC GTC GCT GTC TC - [TAM] 3’

2.5 Site directed mutagenesis PCR

Site directed mutagenesis using PCR was carried out using the Quikchange XL kit (Agilent) using the manufacturers protocol. The primers used for mutagenesis reactions were as follows:

M1 forward primer – 5’-GGG TGT CGA CAT AGC AGA ATA GGC GTT AAT CCA CGA AGG AGA ACA AGA AAT GGA GCC-3’
M1 reverse primer – 5’-GGC TCC ATT TCT TGT TCT CCT TCG TGG ATT AAC GCC TAT TCT GCT ATG TCG ACA CCC-3’
M2 reaction 1 forward – 5’-TCC ATT TCT TGC TCT CCT TTG TGG AGT AAC GCC TAT TCT GC-3’
M2 reaction 1 reverse – 5’-GCA GAA TAG GCG TTA CTC CAC AAA GGA GAG CAA GAA ATG GA-3’
M2 reaction 2 forward – 5’- AGG AGA GCA AGA AAT GGA TCC AGT AGA TCC TAG AC-3’

M2 reaction 2 reverse – 5’-GTC TAG GAT CTA CTG GAT CCA TTT CTT GCT CTC CT-3’
ERK forward – 5’-GAA TAG GCG TTA CTC GAC ATA GGA TAG CAA AAA ATG GAG CCA GTA GAT CC-3’
ERK reverse – 5’-GGA TCT ACT GGC TCC ATT TTT TGC TAT CCT ATG TCG AGT AAC GCC TAT TC-3’

2.5.1 Sequencing

Success of mutagenesis reactions and of cloning was confirmed by sequencing. Sanger sequencing was carried out by an external laboratory (GATC) with appropriate primers on plasmids purified as above.

2.6 Cell culture

2.6.1 Cell lines used

Jurkat cells (ATCC) are a T cell line derived from a patient with T cell lymphoma(Schneider, Schwenk and Bornkamm, 1977). J-lat cells (Aidsreagents) are Jurkat cells transduced with a repressed HIV derived vector that expresses GFP upon stimulation (Chapter 3). SupT1-CCR5 (a gift to our laboratory from

Professor James Hoxie, University of Pennsylvania) is a T cell line that has been stably transduced to express the chemokine receptor CCR5 at a high level. 293T cells (ATCC) are derived from adenoviral transformation of embryonic kidney cells and are stably transfected to express the SV40 virus (simian vacuolating virus 40) large T antigen; this protein permits replication of plasmids containing the SV40 origin of replication and therefore enables increased protein production.

2.6.2 Culture medium

293T cells were cultured in Dulbecco's modified eagle medium (DMEM)(Gibco) supplemented with 10% fetal calf serum (Gibco) and penicillin and streptomycin (Gibco). All additives were filter sterilized using 0.2µm filters.

Jurkat, J-lat and SupT1-CCR5 cells were cultured in Roswell Park Memorial Institute medium (RPMI-1680)(Gibco) supplemented as above.

All cells were cultured at 37°C in the presence of 5% CO_2.

2.6.3 Freezing and thawing cells

When required cells were frozen as 1ml aliquots of $1x10^6$ cells/ml in FCS containing 10% dimethyl sulfoxide (DMSO)(Sigma). Cells were aliquotted into cryovials (Nunc) and stored at -75°C overnight before being transferred to liquid nitrogen where they were stored in the liquid phase.

Cells were thawed by allowing the vial contents to melt at 37°C in a water bath. The cells were then added to 10ml of appropriate medium and transferred to a small tissue culture flask before being placed in the incubator.

2.6.4 Passaging cells

Adherent cells were passaged when they reached 90-100% confluence as judged by microscopy. The medium was aspirated from the flask; cells were washed with sterile phosphate buffered saline to remove any residual serum (PBS, 137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4 \bullet 7H_2O$, 1.4 mM KH_2PO_4, pH 7.3)(Sigma). 1ml of trypsin (Thermofisher) was added to the cells and they were returned to 37°C for

5 minutes. Once the cells were mobilised 10ml of fresh serum containing DMEM was added. 1ml of the resulting suspension was added to a fresh 75cm^2 tissue culture flask with 20ml of DMEM.

Suspension cells were passaged twice weekly. 2ml of suspension was added to a clean tissue culture flask containing 20ml of RPMI. Volumes of media and size of tissue culture vessel were scaled as required for larger numbers of cells.

2.6.5 Counting cells

Cells were counted using a Neubauer chamber. A 10µl aliquot of the cells to be counted was added to 10µl of 0.4% trypan blue (Gibco). 10µl of this was placed into the counting chamber by capillary action at the edge of the cover slip. Live cells do not take up trypan blue. Live cells were counted in one of the 1mm^2 squares that contain a volume of 10^{-4}ml. The concentration of live cells was determined after taking into account the two-fold dilution in trypan blue.

2.6.6 Stimulating cells with latency reversing agents

PMA(Sigma), panobinostat (Sigma), JQ1(Cayman), HMBA (Cayman) and Chaetocin (Sigma) were added to culture medium at the concentrations indicated and cells were cultured in their presence overnight. For experiments involving virus 15nM efavirenz (Aidsreagents) and 100nM raltegravir (Aidsreagents) were added to prevent second round infection or residual virus in the supernatant from infecting stimulated cells.

For two drug stimulation of J-lat cells round bottom 96 well plates were used. Serial two fold dilutions in serum containing RPMI were made of each agent in separate 96 well plates. Each dilution was made 4 times the intended final concentration. 25µl of the first agent was added to each well of the experiment plate working in from top to bottom. 25µl of the second agent was added to each well working from left to right to yield a set of wells with each well containing a different mix of two agents. Finally 10^5 J-lat cells in 50µl serum containing medium was added to each well and mixed.

2.6.7 Spinoculation with virus or vector

Infections and transductions were carried out by spinoculation. Cells and virus containing medium were transferred to a microcentrifuge tube and centrifuged at 800x*g* for 90 minutes at room temperature. Cells were then washed and returned to culture medium in the incubator.

2.6.8 Magnetic bead separation of shRNA transduced cells

Dynabead streptavidin binder beads (Invitrogen) were washed twice in biotin free buffer (Calcium and magnesium free PBS supplemented with 2mM EDTA and 0.1% BSA). Cells were resuspended 10^7 cells/ml in the same buffer and beads were added in a 10:1 ratio of beads to cells. The cell and beads mixture was incubated at 4°C with rotation for 30 minutes. Selection of bead bound cells was carried out by incubating cells in a 5ml round bottom flow cytometry tube in a magnet (Stemcell) for 5 minutes and then washed with biotin free buffer and returned to the magnet. Unbound cells were poured off. Bead bound cells were resuspended in release buffer (RPMI with 10% FCS and 1% penicillin/streptomycin, 10mM HEPES buffer, pH 7.4, 2mM biotin) pre warmed to 37°C and incubated for 15 minutes. Cells were returned to the magnet for 5 minutes and the supernatant containing released cells removed for downstream applications or culture.

2.6.9 Flow cytometry

Flow cytometry was carried out using Accuri (BD) and Attune NXT (Thermofisher) flow cytometers. Cells were suspended in PBS and stained with antibodies or vital stains as required. Cells were washed twice with PBS before being fixed in 4% paraformaldehyde in PBS. Once fixed cells were analysed directly or stored at 4°C in the dark until analysis could take place. Antibodies were used at the concentrations recommended by the manufacturer. DRAQ7 (Abcam) was used to discriminate live and dead cells and was used 1 in 100. Antibodies used were APC conjugated anti-LNGFR (Biolegend) and Vioblue conjugated anti-LNGFR (Miltenyi)

2.7 Preparation of lentiviral vector by calcium phosphate transfection

Pseudotyped HIV viruses were produced by the co-transfection of envelope deficient HIV plasmid with a plasmid expressing vesicular stomatitis virus glycoprotein (VSV-G) into HEK293T cells in a ratio 10:3.5. Vectors for the shRNA transfection were produced using vector plasmids, pCMV-deltaR8.9 and pCMV-VSVG in the ratio 10:5:1.

2.7.1 Large scale preparations using the calcium phosphate method

Reagents were as follows:

2x HEPES buffered saline (BBS) - 50mM HEPES, 10mM KCl, 280mM NaCl, 1.5mM Na_2HPO_2, 12mM dextrose pH 7.05
2M $CaCl_2$

All solutions were sterile filtered using 0.2μm SFCA filters

Plasmids were used at 1mg/ml concentration

Day one

HEK293T cells were seeded onto 20 10cm diameter tissue culture plates (Nunc) in 10ml of medium and cultured overnight

Day two

Medium was changed in the morning of day two. Transfections were set up set up in 5ml bijous, 1 per plate:

500μl 2x HEPES
438μl H_2O
62μl 2M $CaCl_2$
Plasmids in above ratios, total DNA 16μg per plate

The transfection mix was then added drop wise to the plates with swirling to mix.

Day three

Medium was aspirated and replaced with 10ml of fresh serum containing DMEM

Day five

Supernatants from the plates were aspirated using a 50ml syringe then passed through a 0.45µm SFCA filter into six ultracentrifuge tubes (Beckmann). The tubes were then placed into the buckets of a Sorvall Surespin 630 rotor. Supernatants were centrifuged at 25000 rpm at 4°C for 3 hours. Pellets were resuspended in 150µl of 1% BSA and then stored in aliquots of 75µl. Aliquots were frozen at -75°C.

2.7.2 Small scale vector preparations using lipid transfection reagents

Small scale transfections were carried out using the ratios of plasmids above. DNA was mixed with serum free medium and then lipid transfection reagent was added (Turbofect, Thermofisher). Volumes of DNA, medium and transfection reagent were used as per the manufacturer's directions.

2.8 Detection of CA p24 by ELISA

Virus containing supernatants were inactivated by adding the detergent 1% empigen (Sigma) in Tris buffered saline (TBS; 50mM Tris-Cl, 150mM NaCl, pH 7.5) to a final concentration of 0.1%.

High protein binding 96 well plates were coated overnight at room temperature with 25µl per well of HIV-1 p24 antibody diluted in 0.1M bicarbonate solution (25µl coating antibody, 2.25ml water, 0.25ml 1M $NaHCO_3$ - per plate).

The following morning the coating antibody solution was removed by tapping the plates dry on an absorbent surface. Plates were then blocked for 1 hour at room temperature with 100µl per well of 5% bovine serum albumin in TBS. The block was removed prior to the addition of samples and standard. Samples were loaded in duplicate and serially 5x diluted with 0.05% empigen in TBS. A dilution series of p24 standard (Aalto) was prepared using 0.05% empigen in TBS to yield a range of values from 0ng/ml to 10ng/ml. 25µl of sample or standard was loaded in each well. Plates were incubated for 90 minutes at room temperature on a shaker.

The plate was then emptied by tapping onto an absorbent material. Wells were washed four times with 100µl of TBS per well the plate being tapped dry between washes. 25µl of secondary antibody solution was added to each well (alkaline phosphatase conjugated mouse monoclonal anti HIV p24 diluted 1:16000; for one plate 0.05g skim milk powder, 2ml TBS, 0.5ml sheep serum, 12.5µl of 10% Tween 20, 0.15µl antibody). Plates were incubated for an hour at room temperature on a shaker. Plates were washed three times with PBS 0.1% Tween 20 then once with PBS (no Tween 20). 25µl of lumiphos plus (at room temperature) was added to each well and incubated in the dark at room temperature for 30 minutes. Plates were read on a glomax luminometer (Promega).

A linear regression of the standard curve data was used to calibrate measured luminescence to p24 concentration. Sample dilutions that gave results outside the dynamic range of the standard series were discarded. The mean of the two sample repeats was taken.

2.9 Detection of protein by western blot

2.9.1 Lysis of cells

Cell pellets were lysed in radio-immunoprecipitation assay buffer (RIPA; 25mM TRIS, 150mM NaCl, 0.1% sodium dodecyl sulphate, 0.5% sodium deoxycholate, 1% Triton X-100).

2.9.2 Determination of protein concentration

The Pierce BCA assay (Thermofisher) was used to measure protein concentration according to the manufactures protocol. Serial dilutions of BCA standard and sample were loaded into a flat bottomed 96 well plate. Colour substrate was added to each well and mixed before being incubated for 1 hour. Sample absorbance at 590nm was measured using a plate reader (Bio-Rad). Linear regression of the values obtained from the standard was used to determine the protein concentrations in the samples.

2.9.3 Preparation of SDS PAGE gels

PAGE gels were prepared and run using the Bio-Rad mini-protean gel system. 10% acrylamide gels were used for the resolving gel. For 10ml the following was used:

30% acrylamide (37.5:1 acrylamide:bis-acrylamide)3.3ml
1.5M TRIS pH 8.8 2.5ml
10% SDS 0.1ml
H_2O 4ml

5% acrylamide gels were used for the stacking gel. For 10ml the following was used:

30% acrylamide (37.5:1 acrylamide:bis-acrylamide) 1.7ml
1.0M TRIS pH 6.8 1.25ml
10% SDS 0.1ml
H_2O 6.8ml

To initiate polymerisation of the acrylamide 1 in 100 of 10% ammonium persulphate (Sigma) and 1 in 1000 of tetramethylethylenediamine (TEMED, Sigma) were added. Gels were prepared between glass plates using combs of appropriately sized wells. The resolving gel was poured first and permitted to set a few drops of water saturated butanol were added to the top of the gel to ensure a horizontal surface. Once set the butanol was removed and the top of the gel washed twice with water. The stacking gel was then poured in and the comb was then inserted on top. Prepared gels were stored at 4°C wrapped in paper soaked in water.

2.9.4 Loading samples and running gel

Gels were loaded into an electrophoresis tank filled with 1x running buffer (for 1000ml 5x running buffer: 15.1g Tris, 94g glycine, 50ml 10% SDS, 950ml water). 20µg of protein as determined by BCA assay was loaded per well as well as a protein ladder (Precision plus protein ladder, Bio-Rad). Protein extract was mixed 1:1 with loading buffer (10% glycerol, 0.05% bromophenol blue, 1% β-mercaptoethanol, 1% SDS, 62.5mM Tris pH 6.8) and then heated to 95°C for 5

minutes to denature the proteins. Gels were then run at 120V for 2 hours or until the dye in the ladder had separated appropriately.

2.9.5 Transferring gel to membrane

Gels were removed from the glass plates and the stacking gel cut off and discarded. A transfer sandwich of Whatman paper, gel, cellulose membrane (Amersham), Whatman paper was assembled. The paper and membrane were pre soaked in 1x transfer buffer. Transfer buffer was made by diluting 100ml 10x transfer buffer (144g glycine, 30.2g Tris, 1000ml water) with 200ml methanol and 700ml water. Transfer sandwiches were loaded so that the membrane was between the gel and the positive electrode and the tank filled with 1x transfer buffer. Transfers were carried out on ice for 90 minutes at 100V. Membranes were then transferred to blocking solution (5% milk powder in PBS) for at least 30 minutes.

2.9.6 Antibody probing of membranes

Antibodies were diluted in 5ml 5% milk in PBS as the indicated concentration. Membranes were transferred to 50ml centrifuge tubes and incubated overnight with primary antibody at 4°C on a roller. The following morning the membranes were washed 3 times with PBS 0.1% Tween 20 for 10 minutes then incubated for 2 hours with secondary antibody (HRP conjugated-goat anti-rabbit, Santacruz) at 1:2000. The membrane was washed again and then wrapped in film with 0.5ml detection agent (ECL, Amersham). The wrapped films were then exposed to film (Amersham chemiluminescence film).

All primary antibodies used were rabbit IgG.
Anti-Fam208A (Atlas antibodies) used 1 in 5000 (TASOR)
Anti-MPP8 (Proteintech) used 1 in 10000
Anti-GAPDH (Santacruz) used 1 in 20000
Anti-PPHLN1 (Atlas antibodies and Abcam) blots attempted at 1 in 100 to 1 in 5000
Anti0SetDB1 (Abcam) blots attempted 1 in 100 to 1 in 5000

Chapter 3: The effect of combinations of latency reversing agents in the J-lat model of HIV latency

3.1 Introduction

First described in 2003 the J-lat model of HIV latency (Jordan, Bisgrove and Verdin, 2003) has become one of the most widely studied examples of gene silencing in HIV. The model is based on the Jurkat T-lymphocyte cell line. The authors infected Jurkat cell lines with a single round HIV based vector expressing GFP. Two different vectors were used, one which contained a simple LTR-*tat*-IRES-GFP genome and a second which contained a full length HIV genome with an *env* deletion and GFP in place of *nef*. The authors infected Jurkat cells at a low MOI and then sorted cell by flow cytometry to remove the GFP expressing cells. The resulting population contained a mix of cells that were uninfected and cells that had been infected but proviral gene expression had been silenced.

The mixed population of GFP negative cells was stimulated with TNF to induce LTR driven gene expression. The cells that went from negative to positive were sorted and subcloned to produce clonal populations of cells harbouring a repressed provirus. The resulting clones were shown to have little to no expression of GFP in the basal state but could respond to TNF stimulation and make GFP. In their initial paper the authors analysed the integration sites of a small number of the LTR-*tat*-IRES-GFP clones and found that they were integrated into alphoid repeats in the centromeric regions of the chromosome. In contrast a recent study evaluated the integration site of the provirus in four of the full length clones and found that they were integrated into the introns of cellular genes (Symons *et al.*, 2017).

The principal advantages of a cell line based model of HIV latency derive from the fact that all the cells are clonally infected with the same provirus within a single locus. This facilitates comparison between integration events and between

reactivation stimuli. It also enables study of the cellular processes that contribute to silencing in the model. Once of the principal barriers to studying HIV latency using *ex vivo* samples is that latently infected cells are relatively rare in patient samples and among those latently infected cells the number harbouring a replication competent provirus is likely to be smaller still (Ho *et al.*, 2013) thus using a clonal population is one way to overcome this.

A significant downside of the J-lat systems is that by studying a clone at a time it does not recapitulate all the possible ways in which proviruses are silenced *in vivo*. This has been borne out in a comparison of cellular models of HIV latency which included an analysis of J-lat clones in combination with primary cell models and patient samples (Spina *et al.*, 2013), in this analysis although the J-lat cells responded to reversing agents in a manner similar to patient cells, they did not recapitulate the response from patient cells as closely as some of the primary cell models studied.

Despite their shortcomings the J-lat model has been used in many studies and has delivered some useful insights into gene silencing in HIV. It has also been used to study pharmacological approaches to latency reversal that have subsequently gone on to be used in human trials.

It has recently been proposed that clonal proliferation plays a significant part in the maintenance of the latent reservoir (Simonetti *et al.*, 2016; Bui *et al.*, 2017; Hosmane *et al.*, 2017). If this is the case then any optimum strategy to reactivate the latent reservoir not only has to address the question 'is it possible to reactivate proviruses regardless of integration site?' but also 'is it possible to reactivate all proviruses within a clone?'. In the absence of a primary cell model of clonally expanded proviruses or the ability to isolate clones from patient cells the J-lat system is a highly tractable one in which to study optimised activation signals.

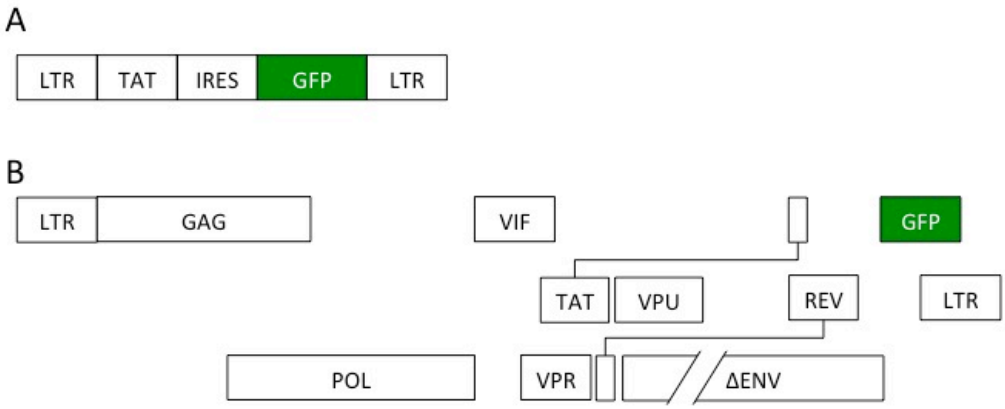


Figure 3.1 Arrangement of proviruses in the J-lat model of HIV latency. **A** truncated provirus present in clones A2 and H2 encoding Tat linked to GFP by an internal ribosome entry site (IRES). **B** Full length provirus present in the numbered clones which expresses GFP in place of Nef and has an envelope deletion to prevent production of infectious virions

J-lat clone	Integration site chromosome	Integration site location	Description of site
6.3	19		Intron of PPP5C
8.4	1	77946384	Intron of FUBP1
9.2	19	46381104	Intron of PPP5C
10.6	9	136468579	Intron of SEC16A
15.4	19	34441293	Intron of UBA2
A2	X		Intron of UTX1

Table 3.1 Published integration sites for the proviruses in the J-lat clones used in these experiments. All clones are integrated into introns. No data was found for the precise integration sites of clone A2 and clone 6.3.

3.2 Experimental approach

The clones selected for this study were A2 (which harbours a truncated LTR-*tat*-internal ribosome entry site (IRES)-GFP provirus) and the full-length clones 15.4, 10.6, 9.2, 8.4 and 6.3. Figure 3.1 shows the arrangement of the genome of the proviruses used in the various clones. These clones were chosen as they were subjected to more in depth analysis in the original paper (Jordan, Bisgrove and Verdin, 2003). Furthermore four of the full length clones have had their precise integration sites defined (Symons *et al.*, 2017) as shown in table 3.1.

Detecting GFP expression from J-lat clones

The GFP expressed by the HIV proviruses in the J-lat cells is readily detected by flow cytometry. Fluorescence can by excited by the 488nm blue laser and emission detected in the visible light spectrum. The experiments were carried out using a Thermo Fisher Attune NXT flow cytometer and the GFP fluorescence was detected using a 530/30 band pass filter. Figure 3.2 shows the gating strategy used to quantify the proportion of cells expressing GFP in response to stimulation.

To exclude dead and dying cells from the analysis cells were also stained with the vital stain DRAQ7. This dye functions as an exclusion stain; it is not taken up by cells with an intact membrane. The dye intercalates cellular DNA and emits in the far-red part of the spectrum. Therefore cells that are DRAQ7 negative are alive and those that fluoresce are dead. The far-red fluorescence was detected using the 695/40 filter of the Attune NXT.

3.3 Determining response to individual latency reversing agents

The latency reversing agents chosen for this study were, panobinostat, JQ1, bryostatin, hexamethylene bisacetamide (HMBA) and chaetocin. The agents were chosen to cover a range of putative latency reversing strategies. Agents

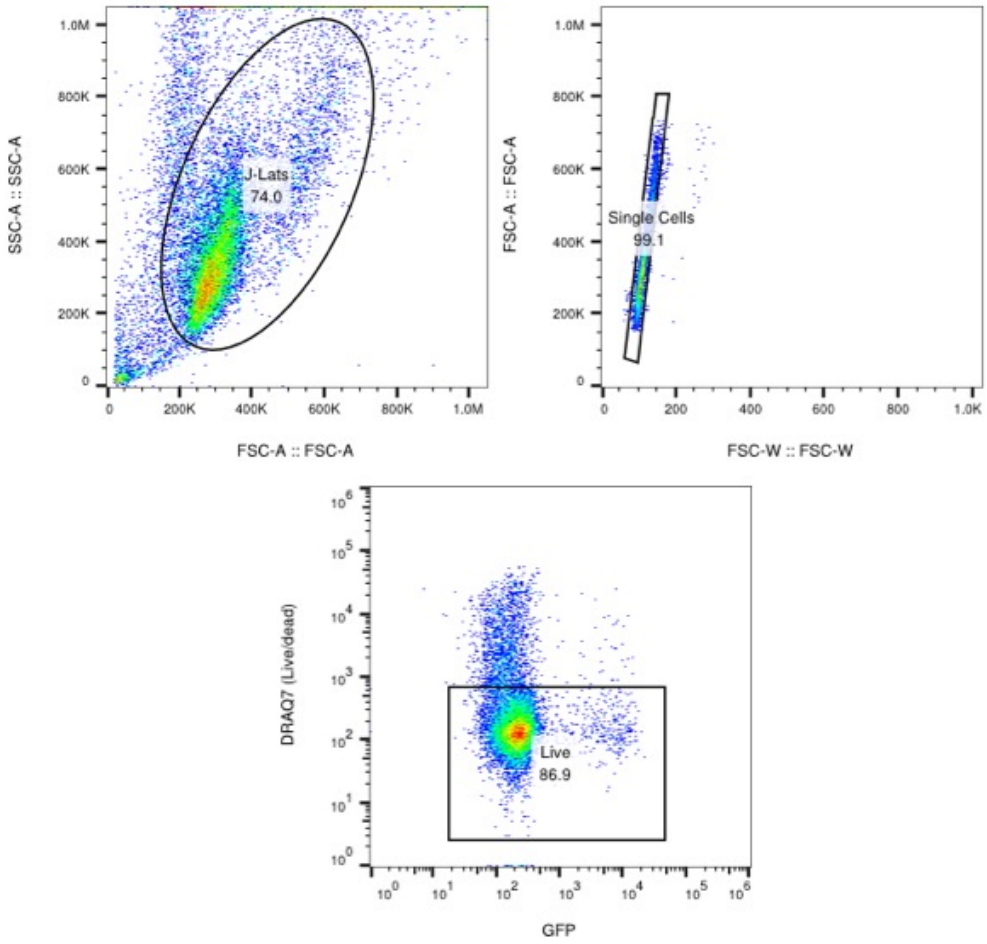


Figure 3.2 Gating strategy employed for the experiments in the chapter. Cells were initially gated on forward (FSC) and side scatter (SSC) characteristics. Single cells were identified by gating forward scatter area (FSC-A) against forward scatter width (FSC-W), single events have a linear relationship between width and area meaning that they form a line, doublets appear as events with twice the expected width for a given area. Live cells were Identified as those which did not take up DRAQ7 stain, an unstained population was used to set the gate. The proportion of GFP expressing cells was determined on the live fraction

were also chosen according to best reflect compounds which have been used *in vivo*, either in the setting of HIV latency or in other clinical settings. The dose ranges for the compounds were taken from the Spina review of model systems of HIV latency(Spina *et al.*, 2013) as well as the papers referenced in the sections below. Phorbol myristate acetate (PMA) was also included in as it has been used as a positive control in previous studies. The maximum dose included in each experiment was at least twice that previously described in the literature in order to ensure a range of relevant doses were covered.

For each experiment cells were seeded into 96 well plates. The medium was replaced with medium containing the latency reversing agent. Agents were used in serial two fold dilutions. GFP expression was measured for each concentration and for each J-lat clone as described in Materials and Methods 24 hours after the addition of the stimulating agent. Data from three experiments were used to generate dose response curves for each agent. By using the logarithmic curve fitting algorithm of the statistical software Prism5 it was possible to derive EC_{50} concentrations for each dose, i.e. the dose required to achieve 50% of the maximum GFP signal.

Dose response to PMA

Figure 3.3 shows the dose response of the six J-lat clones to stimulation with PMA. Each green data point shows the mean and SEM of the %GFP expressing cells for each concentration of PMA. The grey data points show the % of cells gated by forward scatter and side scatter as 'live' which also excluded the DRAQ7 stain. The maximum concentration of PMA used was 400nM. PMA is a potent activator of protein kinase C (PKC) which in turn activates the NFkB pathway leading to activation of lymphocytes. PMA is also used in monocyte lineages to induce differentiation to macrophages. As PMA is an activator of a range of cell types its use *in vivo* is limited as it would cause massive cytokine release. Its role in reversing latent HIV *in vitro* it likely due to an increase in the availability of transcription factors in the activated cell compared to the resting counterpart.

Dose response to panobinostat

Panobinostat is a histone deacetylase inhibitor that acts to boost transcriptions of latent HIV by preventing the removal of the permissive acetyl mark from local histones. It has been trialled *in vivo* (Rasmussen *et al.*, 2014) for the reversal of latent HIV and shown to increase HIV RNA expression and the likelihood of transient HIV viraemia. In addition to the effect on local histones panobinostat may also have a modest effect on cellular activation state (Rasmussen *et al.*, 2013) however the authors found that cellular activation measured by CD69 expression was highly dependent on the level of CD69 expression prior to stimulation. Figure 3.4 shows the dose response of the six clones to panobinostat. The maximum dose of panobinostat used was 800nM.

Dose response to bryostatin

Similarly to PMA bryostatin is an activator of the NFkB pathway through activation of the enzyme PKC. In contrast to PMA Bryostatin has been studied *in vivo;* for the treatment of Alzheimer's disease and lymphoma (Blackhall *et al.*, 2001; Nelson *et al.*, 2017) and is tolerated by human subjects. It should be noted that the plasma levels seen in these trials are 10 to 100 fold lower than *in vitro* doses. Figure 3.5 shows the dose response curves for J-lat clones treated with bryostatin. The maximum dose of bryostatin used was 200nM.

Dose response to JQ1

JQ1 interacts with BET bromodomain containing proteins. These proteins bind PTEFb preventing Tat recruiting it and activating the LTR, JQ1 prevents this binding and it is thought that this increases available PTEFb and increases HIV transcription(Boehm *et al.*, 2013). One study has also identified reduced expression of negative regulators of gene expression on treatment with JQ1(Banerjee *et al.*, 2012). JQ1 itself has a short half life *in vivo* although other BET

inhibitors have made it to clinical trials in the treatment of cancer. Figure 3.6 shows the dose response curves for JQ1, the maximum dose used was 2000nM

Dose response to HMBA

HMBA permits HIV transcription by increasing the pool of PTEFb by releasing it from HEXIM1, a component protein of the 7SK snRNP (Contreras *et al.*, 2007). Figure 3.7 shows the dose response curves for HMBA. The maximum dose used was 10mM.

Dose response to chaetocin

Chaetocin is a broad inhibitor of histone lysine methyltransferases (HMTs) which place repressive chromatin marks(Cherblanc *et al.*, 2013). It is proposed as a latency reversing agent in that by stopping the deposition of repressive methyl marks it will permit expression from latent HIV proviruses. In *ex vivo* studies of its effect on HIV transcription it has been shown to have a modest effect on HIV mRNA transcription(Bouchat *et al.*, 2012). Figure 3.8 shows the dose response curves. The maximum dose of chaetocin used was 800nM

Summary of single agent stimulations

These data show a number of interesting points. Firstly, each clone appears to have a fixed upper proportion of cells that can be made to express GFP. For example approximately 30% of clone 9.2 cells produced GFP in response to maximum PMA stimulation in contrast to approximately 90% for clone 10.6. Additionally beyond a certain concentration of stimulating agent the proportion of responding cells does not increase. This suggests the difference in max %GFP is not simply a difference in dose response between each clone. Similar maximum %GFP were seen using the different agents within a clone i.e. a clone which gives a low max %GFP with one agent could be expected to produce a low max %GFP with other agents.

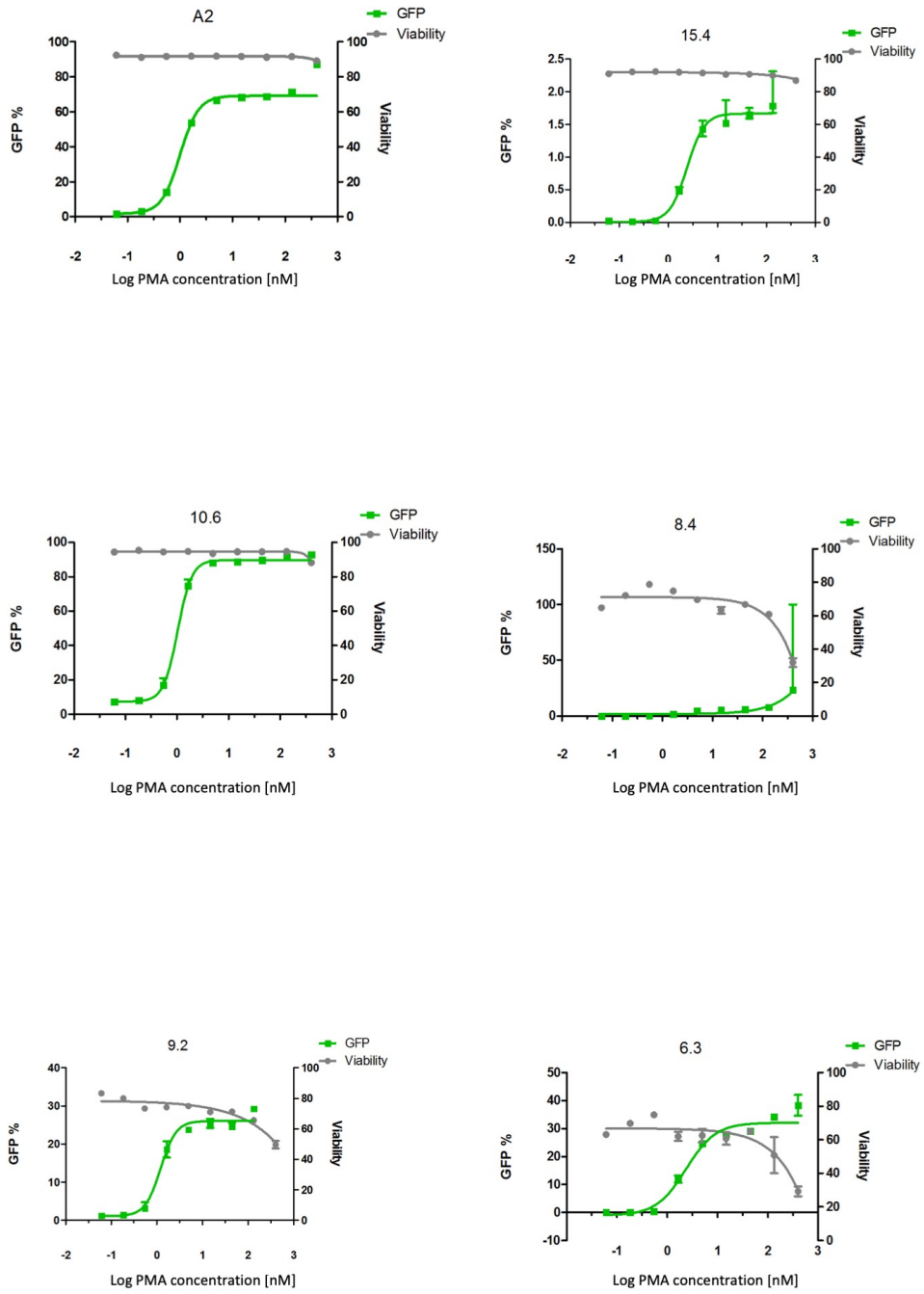


Figure 3.3 – Dose response of J-lat clones to stimulation with PMA. Green points show mean and standard error of % GFP expressing cells for three experiments (left axis, axis scaled to maximum GFP). Grey points represent the % cells gated 'live' by DRAQ7 exclusion (right axis). Lines represent curves fit by logarithmic curve fitting.

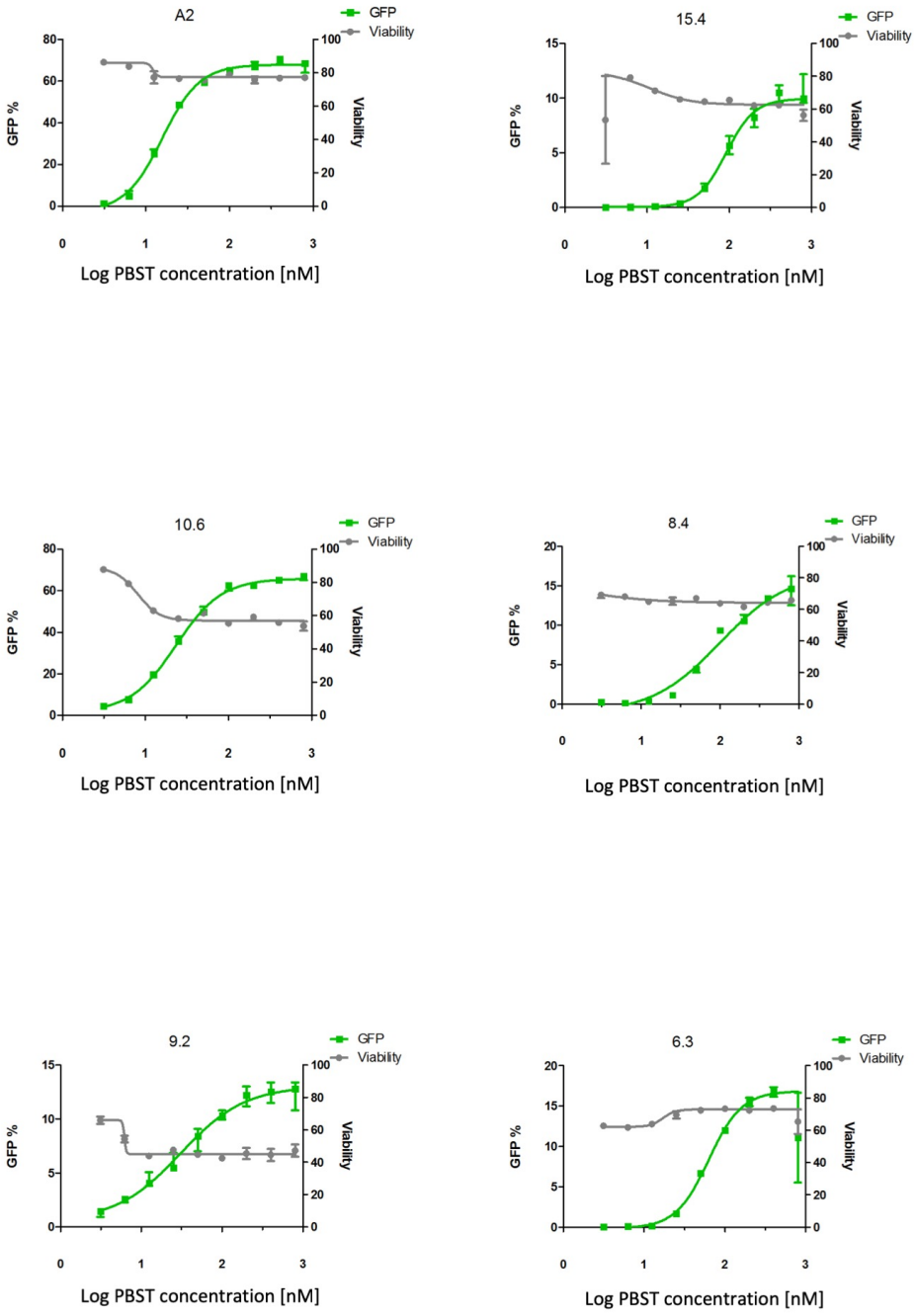


Figure 3.4 – Dose response of J-lat clones to stimulation with panobinostat. Green points show mean and standard error of % GFP expressing cells for three experiments (left axis, axis scaled to maximum GFP). Grey points represent the % cells gated 'live' by DRAQ7 exclusion (right axis). Lines represent curves fit by logarithmic curve fitting.

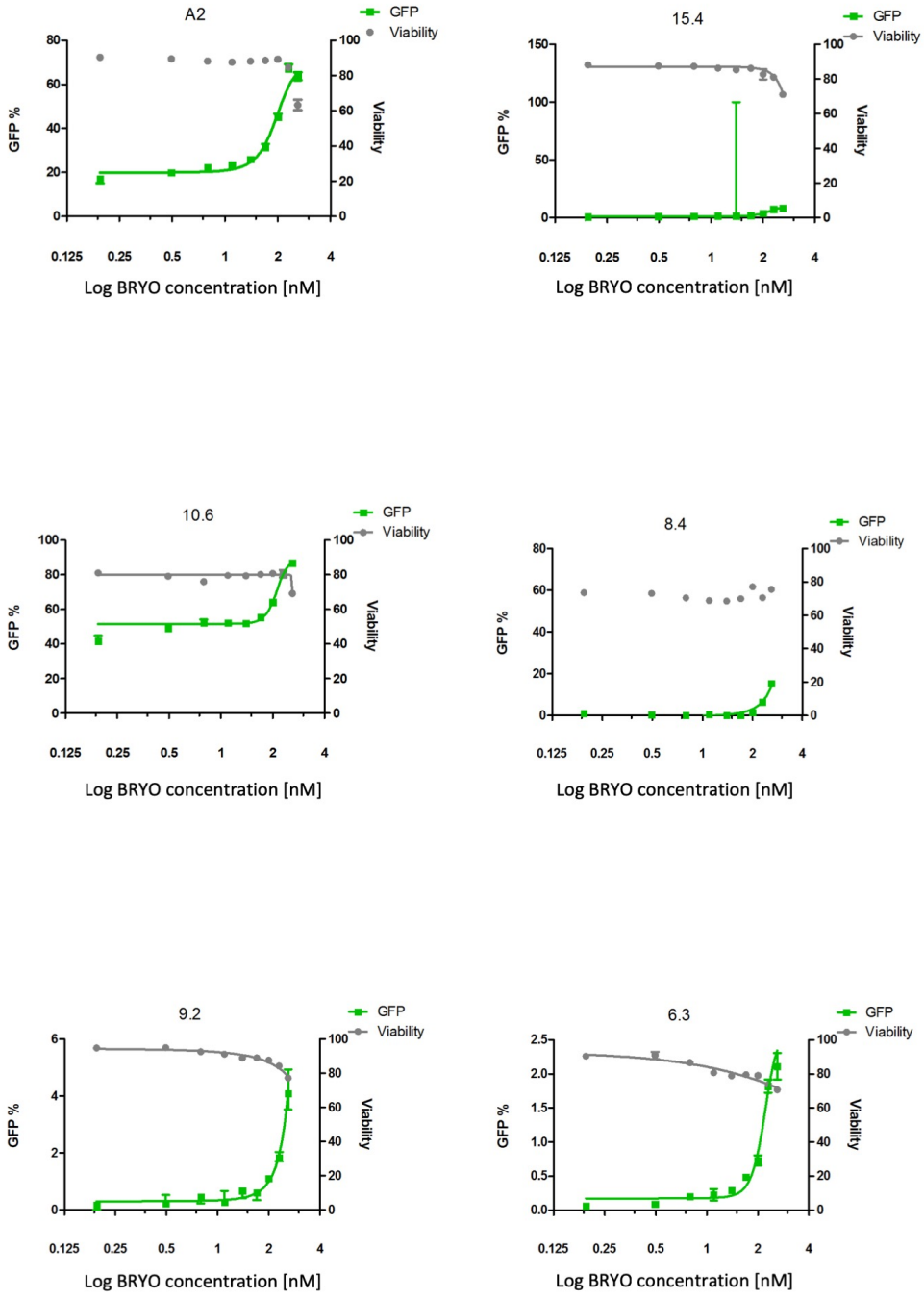


Figure 3.5 – Dose response of J-lat clones to stimulation with Bryostatin. Green points show mean and standard error of % GFP expressing cells for three experiments (left axis, axis scaled to maximum GFP). Grey points represent the % cells gated 'live' by DRAQ7 exclusion (right axis). Lines represent curves fit by logarithmic curve fitting. Where no line is shown it was not possible to fit a curve to the data

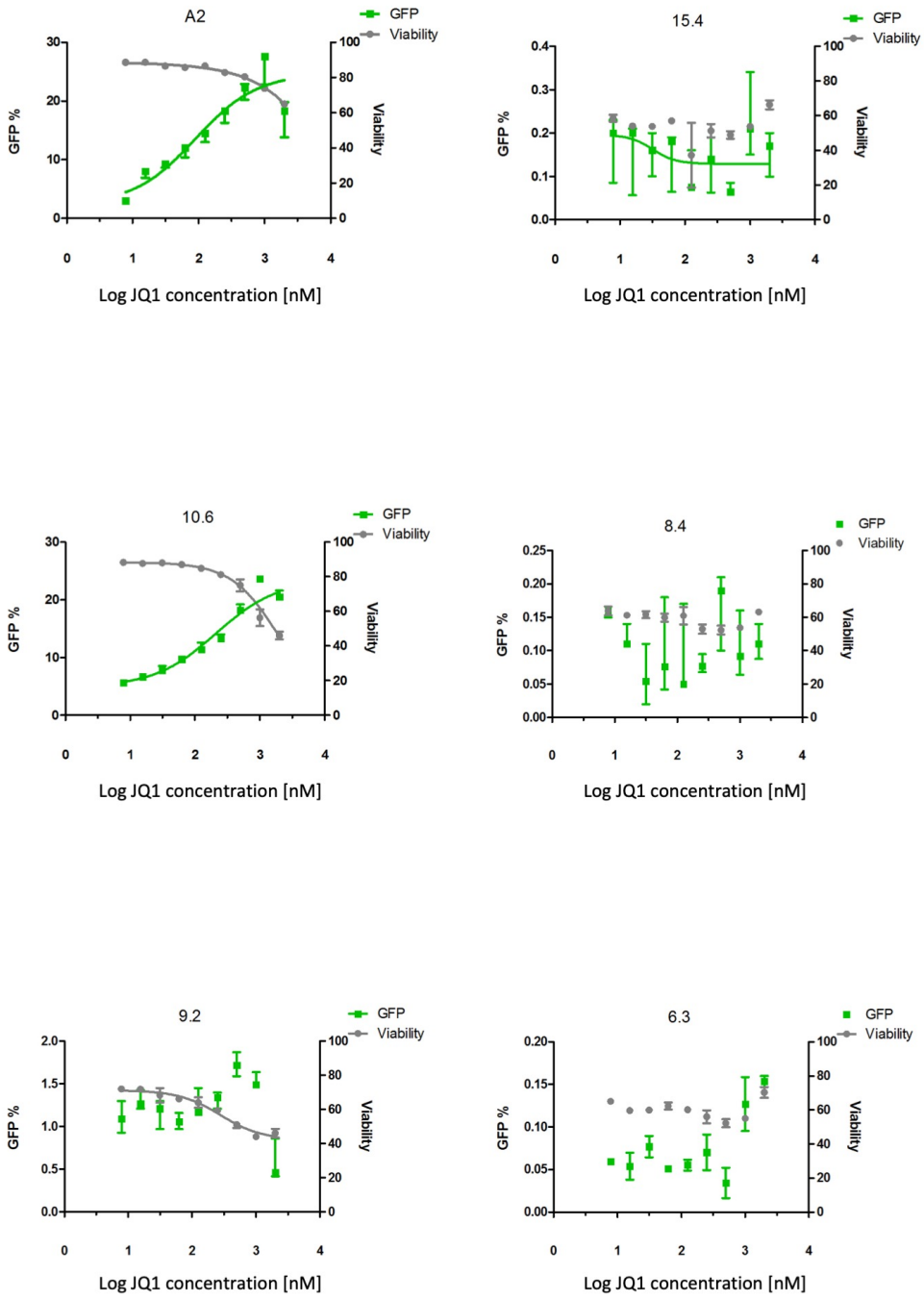


Figure 3.6 – Dose response of J-lat clones to stimulation with JQ1. Green points show mean and standard error of % GFP expressing cells for three experiments (left axis, axis scaled to maximum GFP). Grey points represent the % cells gated 'live' by DRAQ7 exclusion (right axis). Lines represent curves fit by logarithmic curve fitting. Where no line is shown it was not possible to fit a curve to the data

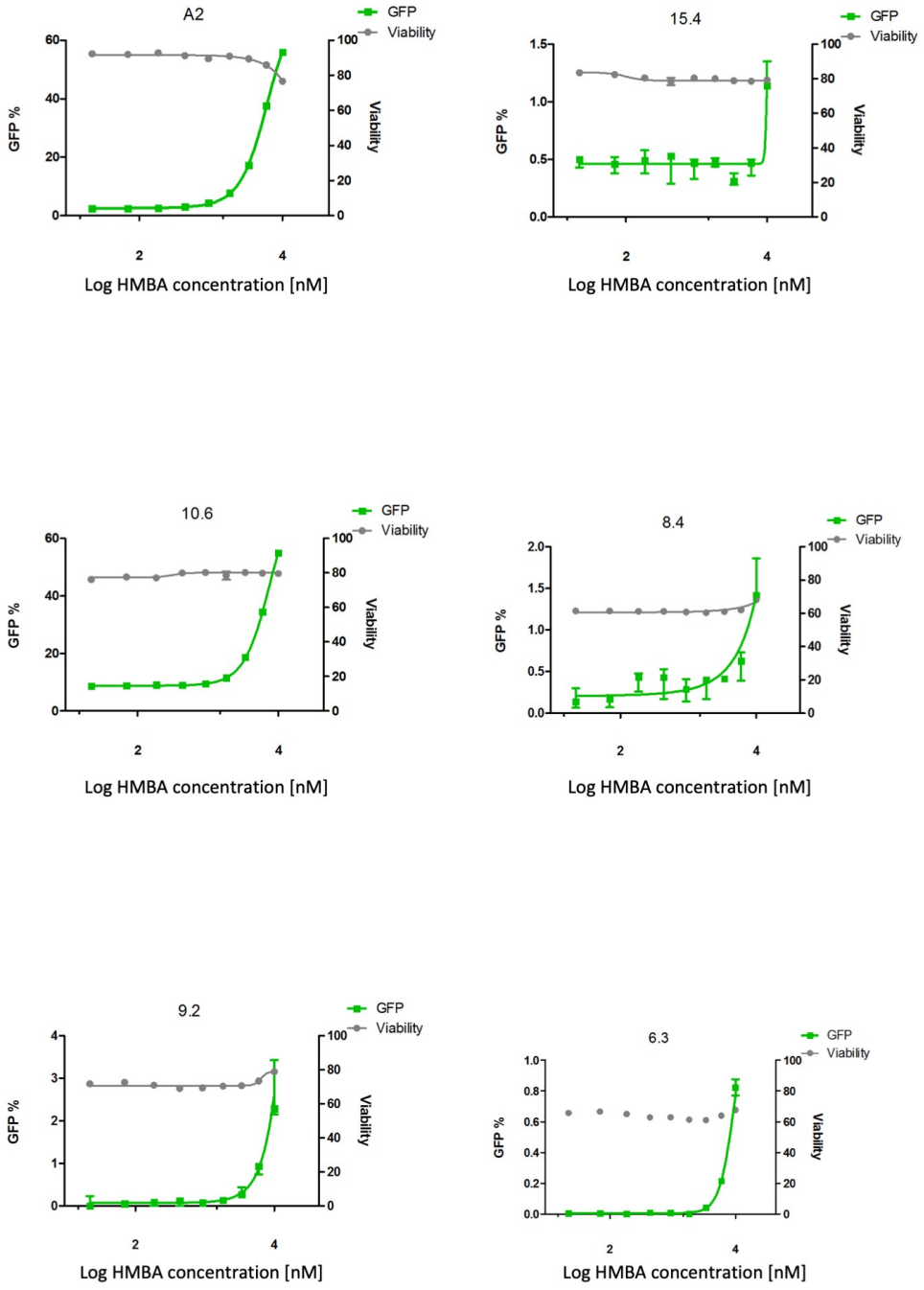


Figure 3.7 – Dose response of J-lat clones to stimulation with HMBA. Green points show mean and standard error of % GFP expressing cells for three experiments (left axis, axis scaled to maximum GFP). Grey points represent the % cells gated 'live' by DRAQ7 exclusion (right axis). Lines represent curves fit by logarithmic curve fitting. Where no line is shown it was not possible to fit a curve to the data

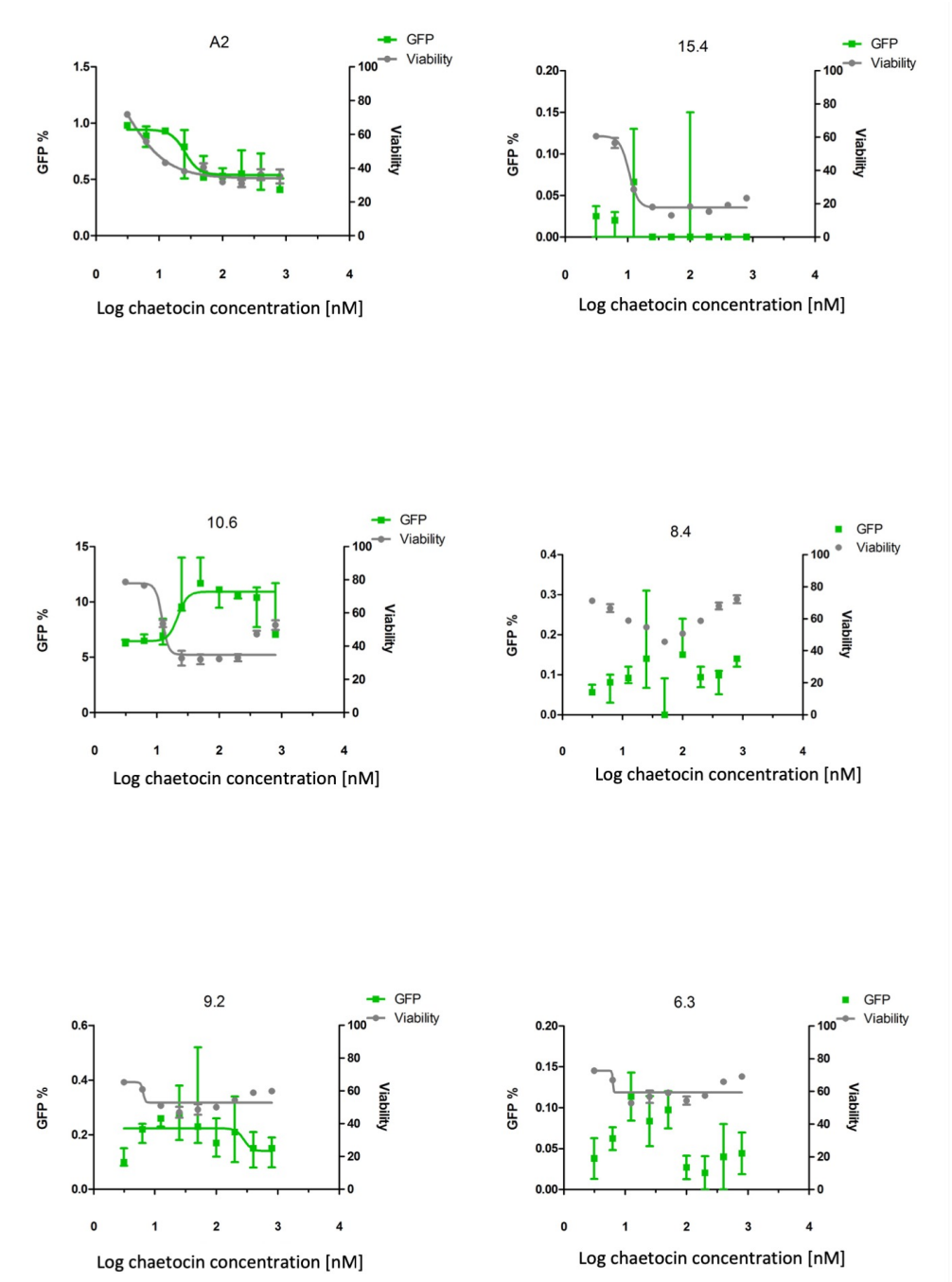


Figure 3.8 – Dose response of J-lat clones to stimulation with chaetocin. Green points show mean and standard error of % GFP expressing cells for three experiments (left axis, axis scaled to maximum GFP). Grey points represent the % cells gated 'live' by DRAQ7 exclusion (right axis). Lines represent curves fit by logarithmic curve fitting. Where no line is shown it was not possible to fit a curve to the data

J-lat clone	PMA[nM]	PBST[nM]	BRYO [nM]	JQ1[nM]	HMBA [µM]	Chaetocin [nM]
A2	1.002	15.7	88.7	89.3	4798	25
15.4	2.391	91.14	135	-	9943	-
10.6	1.036	23.4	125.7	212	6188	21.7
9.2	1.2	30.7	-	-	-	-
8.4	3.9	99.45	243	-	-	-
6.3	2.35	62.88	151.4	-	9011	-

Table 3.2 – EC_{50} obtained by algorithmic curve fitting for each agent in each clone. EC_{50} was defined as the dose required to achieve 50% of the maximum %GFP observed for each combination of stimulus and clone. Missing data point represent data sets where it was not possible for an EC_{50} to be calculated. The EC_{50} PMA was similar for all clones however larger differences were seen with other activating agents.

Secondly the shape of the dose response curves to PMA are similar for all of the clones. This suggests that proviruses have a similar response to a broad cellular activator regardless of integration site. Table 3.2 shows the EC_{50} values for each activating agent and each clone for which the data permitted a curve to be fitted. Clones 15.4 and 8.4 had higher EC_{50} to panobinostat suggesting the integration site may be less responsive to manipulation of histone acetylation then in other clones. The response to JQ1 was only consistent in clones 10.6 and A2; other clones did not appear to respond to JQ1 suggesting that some integration sites may not be responsive to BRD inhibitors. Also 10.6 and A2 appear to be the clones that give the largest increase in GFP signal for all activating agents and it is possible that the activation signal was too small to reliably measure for JQ1 in the other clones. Similarly clone 9.2 gave inconsistent results with the agents HMBA and bryostatin meaning the curve fitting software was unable to derive an EC_{50}.

Thirdly, none of the clones gives a clear activation response to chaetocin. This is in contrast to the studies mentioned above however it should be noted that the authors of those studies measured p24 production and mRNA levels rather than proportion of cells activated by chaetocin. In the data here chaetocin is clearly toxic to the cells at the higher end of the dose range. It is possible that mRNA or p24 production was increased however GFP expression was limited by cell death.

3.4 Stimulation of J-lat clones with panobinostat in combination with other latency reversing agents

As described above the clones each have set maximal proportions that respond to stimulation and these differ for each activating agent. To evaluate whether stimulation in combination could increase the proportion of cells that responded to levels above this maximum the cells were stimulation in combination. As panobinostat gave the most robust activation signal for all clones it was chosen as the base reversal agent to which other agents were added. To enable evaluation over a range of concentrations of panobinostat and the second activating agent the cells were stimulated in a grid of different concentrations of agents. For each

experiment the J-lat cells were seeded into 64 wells of a 96 well plate. 7 serial two fold dilutions of panobinostat were placed into rows of the plate with a final eighth row containing medium only. Then into the columns of the plate a second agent was added, again in serial two fold dilutions with a final row containing medium only. Therefore for each plate there was one unstimulated well, one column of wells only stimulated with dilutions of panobinostat and one row only stimulated with the second activating agent. There were 49 wells stimulated with both activating agents at various ratios. As previously the cells were stained with DRAQ7 to permit exclusion of dead cells and measurement of GFP expression from living cells only.

Figures 3.9 to 3.13 show the data for the different combinations. Each figure shows the GFP expression data in graphical form. The top row (**A**) in each figure shows the raw % GFP expression colour coded to show the percentage, with 100% being red and 0% blue.

The second row(**B**) shows the same numerical data however the squares are coloured to show the highest value obtained as the darkest green with other values scaled accordingly. This was done to make it easier to appreciate the similarities between clones even when the numerical maximum differed between them.

The effect of combinations may be additive or synergistic; synergistic effects are those where the effect of two compounds is together is greater than the two in addition. Synergies imply that the effect of one compound enhances the effect of the other. To determine whether the observed effect of a combination was synergistic a score was derived; this score was based on the Bliss model for predicting the additive effect of two drugs. The Bliss model was used as simple addition of two percentages does not yield useful results, consider two drugs which each activate 60%, simply adding them together suggests they would activate 120% which is clearly not possible.

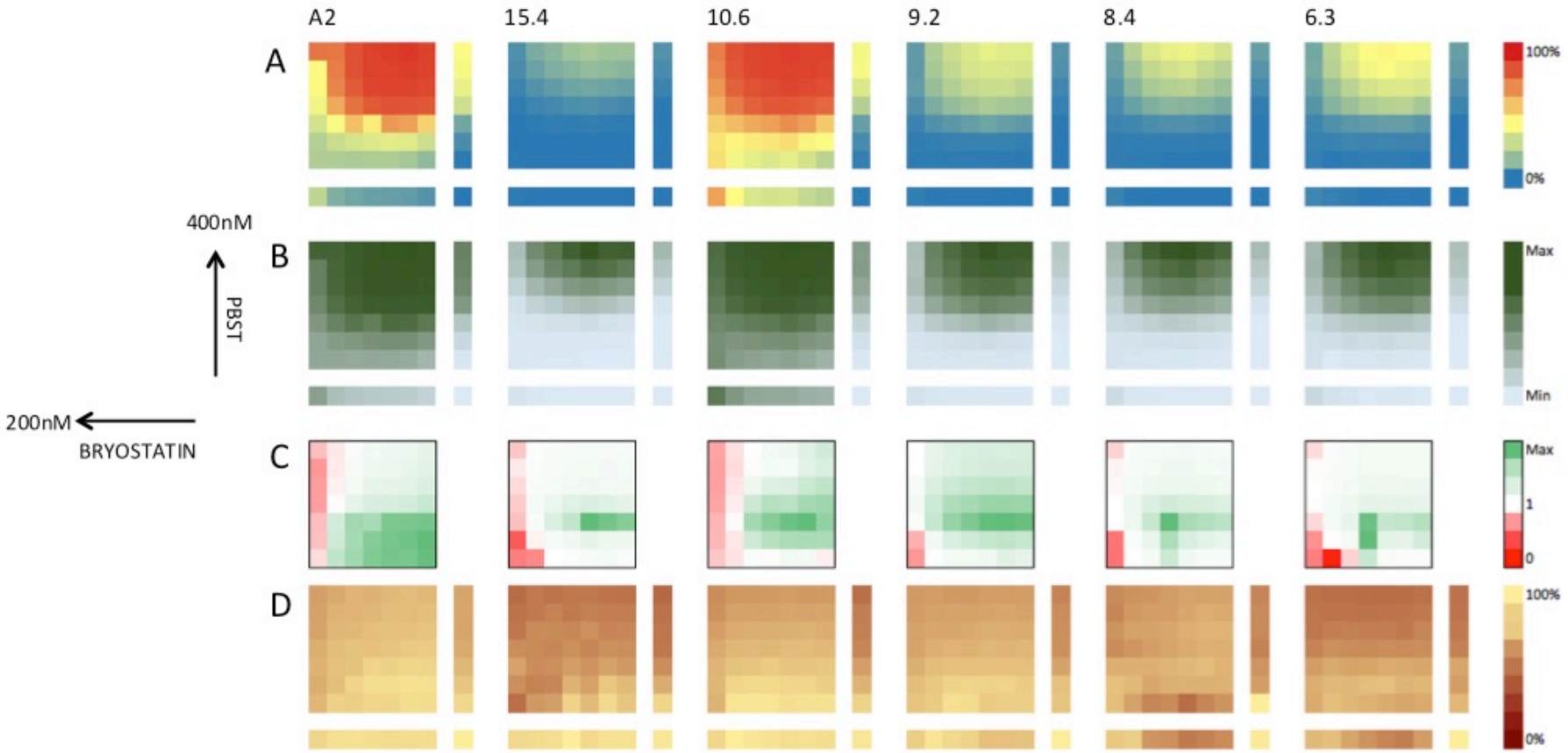


Figure 3.9 – Effect of bryostatin in combination with panobinostat on GFP expression from J-lat clones. **A** shows the raw GFP expression in percent. **B** shows the expression normalised to maximum observed GFP signal. **C** shows the calculated synergy scores for each combination. **D** shows the relative viability of cells compared to the unstimulated sample. Single rows and columns at the right and lower edges show the effect of single agents. Color keys are given at the right side of the figure. The arrows at the right show the maximum concentration of each agent and the direction of increasing concentration of each agent.

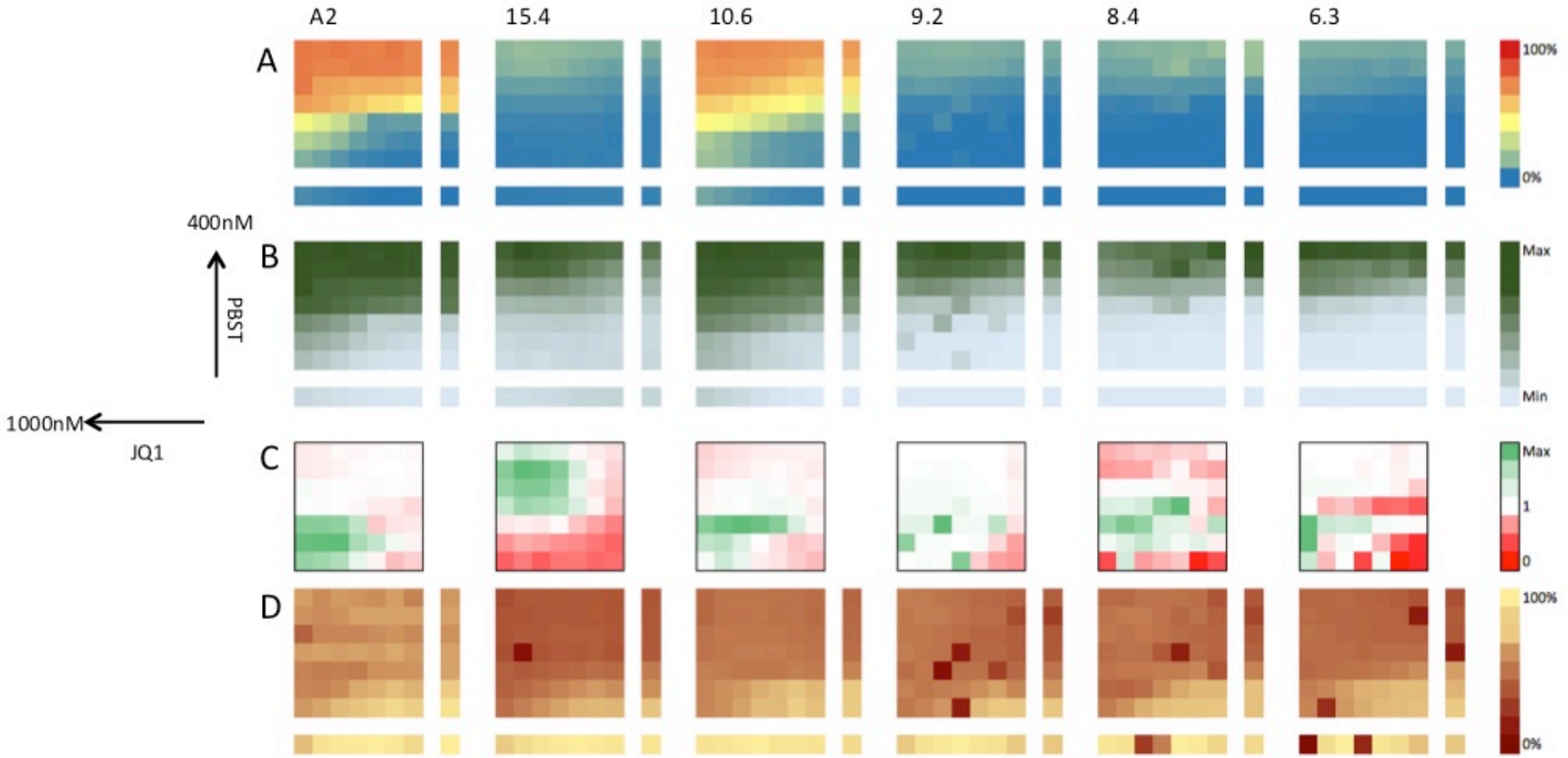


Figure 3.10 – Effect of JQ1 in combination with panobinostat on GFP expression from J-lat clones. **A** shows the raw GFP expression in percent. **B** shows the expression normalised to maximum observed GFP signal. **C** shows the calculated synergy scores for each combination. **D** shows the relative viability of cells compared to the unstimulated sample. Single rows and columns at the right and lower edges show the effect of single agents. Color keys are given at the right side of the figure. The arrows at the right show the maximum concentration of each agent and the direction of increasing concentration of each agent.

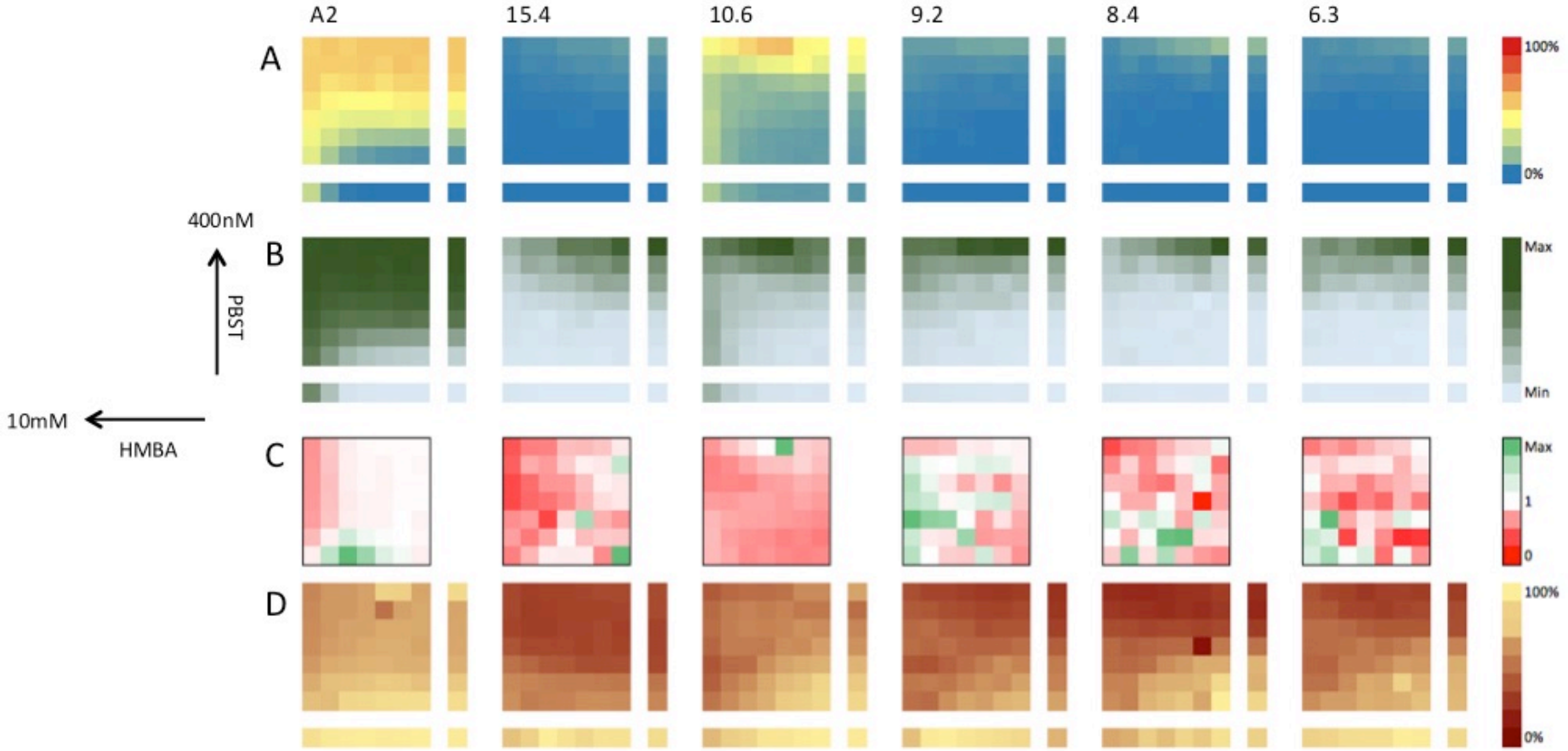


Figure 3.11 – Effect of HMBA in combination with panobinostat on GFP expression from J-lat clones. **A** shows the raw GFP expression in percent. **B** shows the expression normalised to maximum observed GFP signal. **C** shows the calculated synergy scores for each combination. **D** shows the relative viability of cells compared to the unstimulated sample. Single rows and columns at the right and lower edges show the effect of single agents. Color keys are given at the right side of the figure. The arrows at the right show the maximum concentration of each agent and the direction of increasing concentration of each agent.

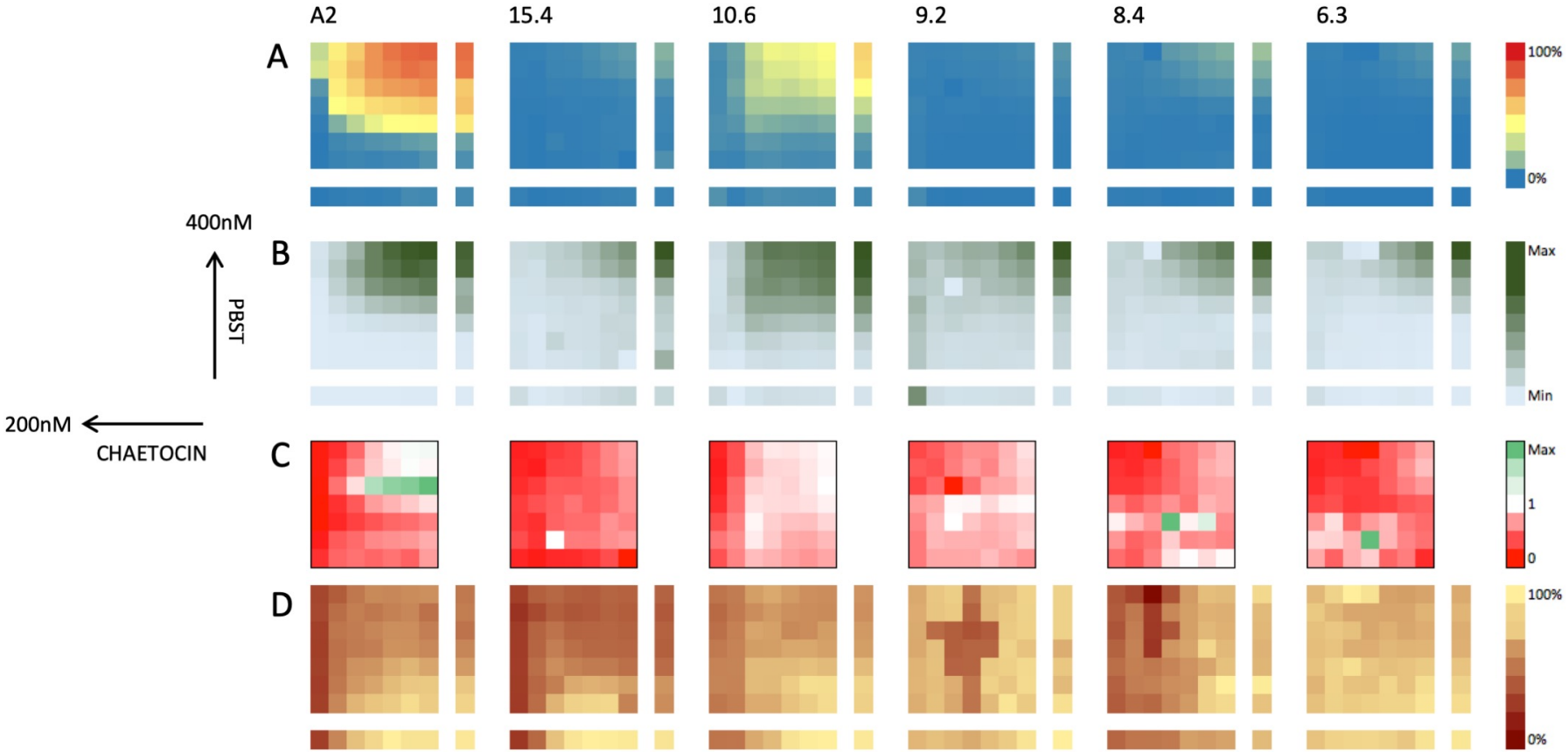


Figure 3.12 – Effect of chaetocin in combination with panobinostat on GFP expression from J-lat clones. **A** shows the raw GFP expression in percent. **B** shows the expression normalised to maximum observed GFP signal. **C** shows the calculated synergy scores for each combination. **D** shows the relative viability of cells compared to the unstimulated sample. Single rows and columns at the right and lower edges show the effect of single agents. Color keys are given at the right side of the figure. The arrows at the right show the maximum concentration of each agent and the direction of increasing concentration of each agent.

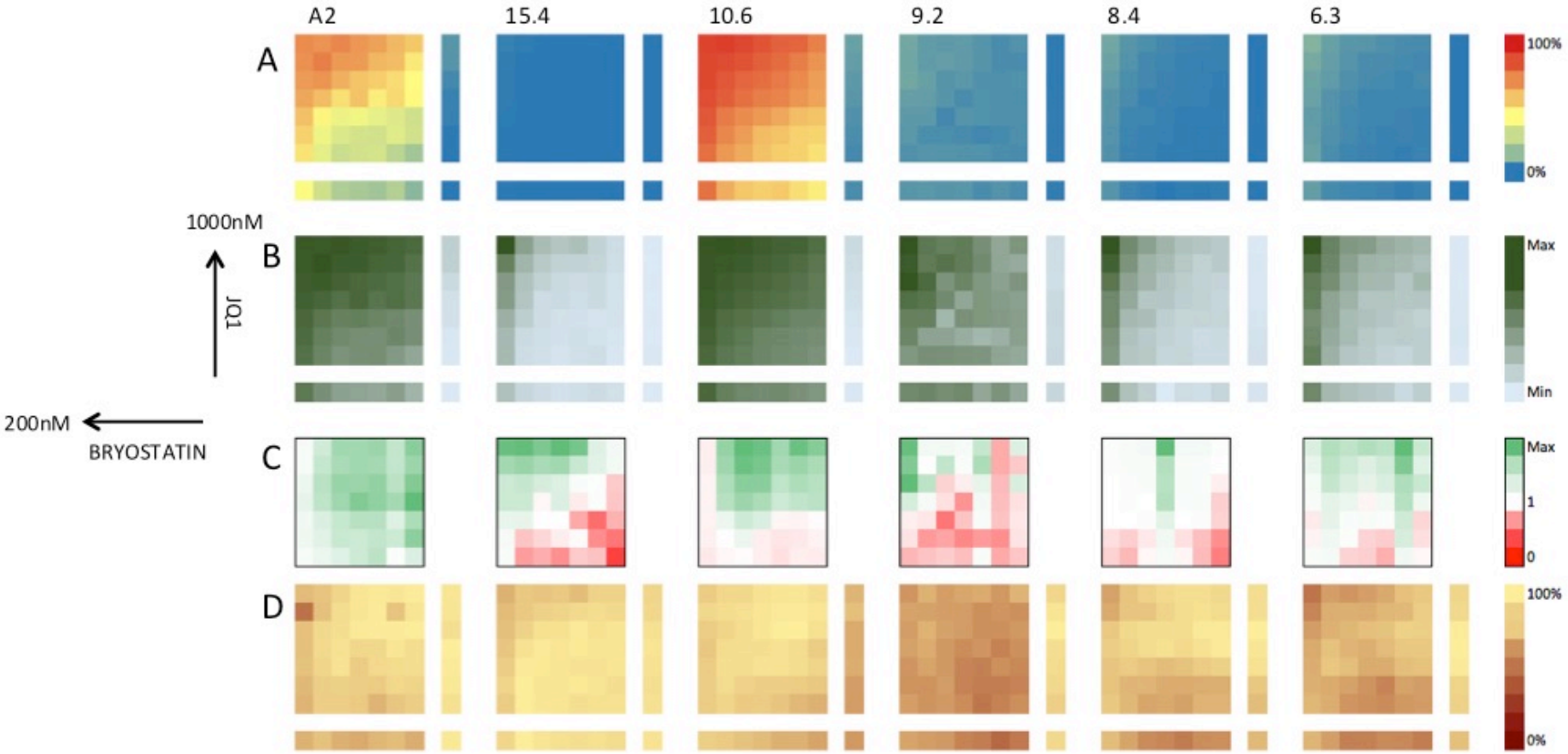


Figure 3.13– Effect of bryostatin in combination with JQ1 on GFP expression from J-lat clones. **A** shows the raw GFP expression in percent. **B** shows the expression normalised to maximum observed GFP signal. **C** shows the calculated synergy scores for each combination. **D** shows the relative viability of cells compared to the unstimulated sample. Single rows and columns at the right and lower edges show the effect of single agents. Color keys are given at the right side of the figure. The arrows at the right show the maximum concentration of each agent and the direction of increasing concentration of each agent.

common pathway. In this case the readout is through LTR mediated expression of GFP. The Bliss model is that the additive effect of two drugs A and B is the proportional effect of A plus the proportional effect of B on those not activated by A; which is given by the equation

$$P_{AB} = P_A + P_B(1 - P_A)$$

For the example above of two drugs with an effect of 60% the Bliss model predicts that added together they would have an effect of 84% (60% plus 60% of the remaining 40%). To derive a score for this the observed %GFP for each combination was divided by the %GFP predicted by the Bliss model from the single agent data. Scores around 1 suggest that the effect is additive, scores greater than 1 suggest the effect is synergistic and scores of less than one suggest than the effect is negatively synergistic.

The third row(**C**) in each of the figures shows the synergy score. Scores less than one are coloured red, scores of 1 are coloured white and scores greater than one are green.

The final row (**D**) in the figures shows the effect of the agents in combination on viability. The percentage live cells in the medium only well was used as a control and the proportional viability seen in other wells scaled accordingly. Darker coloured boxes indicate reduced viability compared to the control well. Viability for each well was derived as described above in the single agent experiments

Stimulation of J-lat clones with Bryostatin and JQ1

In addition to the combinations with panobinostat the clones were also stimulated with bryostatin and JQ1 in combination (Figure 3.13). This was done to evaluate whether synergy occurs in drug combinations other than with panobinostat and also as they have been identified as a potent combination for activating HIV transcription in patient cells examined *ex vivo*(Darcis *et al.*, 2015). These two agents were used as they act via two distinct pathways, bryostatin through

increasing the availability of transcription factors and JQ1 through increasing the available PTEFb. It is reasonable to hypothesise that more synergy might be be observed in combinations that affect two distinct pathways than with combinations both acting via the same pathway. Conversely a combination of agents acting via the same pathway such as HMBA and JQ1 both act through PTEFb and therefore were predicted not to yield synergy.

3.5 Discussion

The single agent stimulations show that, despite being clonal, cells with an identical vector integration site do not all respond to stimulation. For example only approximately 30% of cells from clone 9.2 produced GFP in response to PMA stimulation. This is unexpected as one would predict that as the cells are identical, the provirus is in the same genetic context and each cell is exposed to the same concentration of activating agent each provirus should have an equal response to activation. In fact the observed data suggest something different which is that the maximum proportion of proviruses from a clonally expanded integration site which respond to stimulation is limited by additional factors affecting the integration site. In the clones harbouring 'full length' proviruses the integration site accounts for a large range of response rates to PMA stimulation from 3% (clone 15.4) to 90% (clone 10.6). The other important observation here is that integration sites also affect the threshold dose required to activate the cells, for example there is a 6.3 fold difference in the concentration of panobinostat required to achieve a 50% signal in clone 8.4 compared to A2.

This has important implications for HIV cure. Clonally expanded populations of proviruses within patient reservoirs have been reported(Bui *et al.*, 2017; Lee *et al.*, 2017). The proportion of the latent reservoir made up of clones is not yet known. If multiple clones in different integration sites are present within an individual the data here suggest that the clones could have variable response rates to a single round of stimulation and thus would decay at different rates dependant on the proportion of the clone activated at each stimulation. For example if we consider

the *in vivo* situation for two clones, one clone which has a 5% response rate to a given latency reversing agent and another with a 30% response rate, it would take 14 rounds of treatment to have activated 50% of the first clone versus only two rounds for the second. Thus we could expect that within patients there will be clones that show variable 'resistance' to a kick and kill strategy for elimination.

Therefore the best strategy for eliminating a clonally expanded populated of latently infected cells is to maximise the percentage of cells which respond to treatment. The data here suggest that combinations of latency reversing agents will be essential to maximise reactivation. The most promising combination studied here was panobinostat and bryostatin. These two agents had a clear synergy and could significantly boost the proportion of cells responding to stimulation. For example in here in clone 9.2 the maximum responses observed with bryostatin or panobinostat alone were 2.3% and 10.1% respectively. In contrast the maximum response for the two agents in combination was 45% observed for panobinostat 400nM with 25nM bryostatin. The synergistic effects of this combination in a similar model system of HIV latency have been described (Martínez-Bonet *et al.*, 2015) although these authors studied a limited range of concentrations in two clones.

The maximum Bliss scores were highest for clones that gave the smallest response to individual agents. This is to be expected, consider two agents that individually give a 90% response rate, an additive effect alone would yield a response of 99%. The experimental method used here means it may not be possible to detect response rates that high as, despite gating for live and dead cells, it is impossible to eliminate cell debris that gives a false 'no GFP' signal. The maximum Bliss score observed was 36.8 meaning that the observed GFP signal was 36.8 times greater than would be predicted if the two agents had an additive effect alone.

The point of maximum synergies (as calculated using the Bliss score given above) for the two agents were also fairly consistent between clones and was seen at 25nM panobinostat with 25nM bryostatin or an adjacent combination when averaged and had an average value of 15.68 meaning the observed GFP signal was

15 times what would be observed if the effect was additive the highest average for any of the combinations tested. Other areas of synergy seen were high concentrations of JQ1 with low concentrations of panobinostat and high doses of JQ1 with intermediate doses of bryostatin. Negative synergy scores were seen across all concentrations for chaetocin and panobinostat suggesting one of the two agents may directly antagonise the effect of the other.

The data here suggest that the optimum strategy for combining HDAC inhibitors with other latency reversing agents to maximise the proportion of cells which respond is to combine panobinostat and bryostatin. One particularly important aspect is the finding that this combination is most effectively synergistic at low concentrations. This suggests that combination dosing may achieve useful synergies at doses that are present *in vivo*. Future studies to deplete the reservoir could make use of this finding to rationally reduce drug toxicities while still usefully affecting the reservoir.

Chapter 4: Investigating the role of HUSH mediated recruitment of SETDB1 in HIV latency

4.1 Introduction

The Human Silencing Hub (HUSH) complex is a recently identified group of proteins first described by the Lehner group at the University of Cambridge in their landmark paper (Tchasovnikarova *et al.*, 2015). In the same paper it was shown that HUSH has at least three components (MPP8, PPPHLN1 and TASOR) and functions to recruit the histone lysine methyltransferase SETDB1 to areas of heterochromatin to catalyse the deposition of H3K9 trimethylation. In their initial screening experiment the near haploid cell line KBM7 was transduced with a GFP reporter driven by the spleen focus forming virus (SFFV) promoter. Cells were sorted according to the level of GFP expression into dim and bright populations. Cells from the dim population were then transduced with a gene trap vector that causes loss of function in cellular genes by the insertion of a promoterless gene product with a 3' splice site. The gene trap vectors also delivered an mCherry transduction signal. Sorting those cells that were expressing mCherry and had moved from GFP-dim to GFP-bright enabled an unbiased screen for genes that were mediating the repression of the SFFV promoter.

Mapping the integration sites of the gene traps that caused an increase in GFP expression in the screen identified SETDB1, MPP8, PPHLN1 and TASOR. SETDB1 (also known as ESET) was already known as a histone lysine methyltransferase, MPP8 had previously been shown to bind to H3K9me3 and to interact with SETDB1 (Kokura *et al.*, 2010). PPHLN1 has been found to be involved in cell cycling; over expression results in the repression of cellular genes(Kurita *et al.*, 2007). The gene which was renamed TASOR by the Lehner group had previously been designated FAM208A and did not have any functions ascribed to it. The model proposed by the Lehner group is that the HUSH complex identifies regions of H3K9me3 and recruits SETDB1 to deposit further repressive methyl marks to

the chromatin. They demonstrated that CRISPR mediated knockdown of the HUSH components resulted in loss of H3K9me3 at a wide range of genomic loci.

The effect of HUSH on HIV promoters was also studied. KBM7 cells were transduced with an LTR-*tat*-IRES-GFP construct. Poorly expressing (GFP-dim) cells were sorted and HUSH components were knocked down by shRNA. The GFP dim cells became bright on knockdown of the HUSH components. The group also studied the effect of knockdown in four J-lat clones harbouring a silenced LTR-*tat*-IRES-GFP construct and observed an increase in the percentage of GFP expressing cells in three of the four clones tested. The strongest effect they observed was in the H2 clone (from 10% GFP in the shControl group to 20% in the HUSH knockdowns); one clone that did not respond was clone A2.

SETDB1 has previously been identified as having a role in the control of HIV transcription. SETDB1 has been found to methylate HIV Tat protein at lysine 51 and this reduces its ability to transactivate the LTR (Van Duyne *et al.*, 2008). The authors of this paper also demonstrated that depletion of SETDB1 increased expression from the LTR of both integrated and transient expression systems.

Another study to have proposed a link between SETDB1 and HIV identified SETDB1 as a possible restriction factor for HIV(Liu *et al.*, 2011). This group used a genome wide library of siRNAs that were transfected into HeLa cells. The cells were then infected with a single round HIV vector. In this study knockdown of SETDB1 resulted in increased expression of HIV proteins, interestingly the study authors reported an increase in integrated HIV DNA in the context of SETDB1 knockdown suggesting an inhibitory effect at some step prior to integration that was relieved by knocking down the protein.

In contrast to the above studies a screen of shRNA knockdown in two latently infected cell lines containing more complete HIV proviruses (Nguyen *et al.*, 2017) did not identify SETDB1 as a repressive factor. This study utilised Jurkat cells, a more physiologically relevant cell line for models of HIV infection, which were infected with a delta-Gag provirus expressing d2EGFP in place of Nef. The study

was designed specifically to identify histone lysine methyltransferases responsible for maintaining silencing in these model systems. Although SETDB1 was included in the shRNA library it was not found to be a 'hit'. The authors found a role for members of polycomb repressor 2 and in particular EHZ2 in maintaining latency but not for other histone lysine methyltransferases. Another study which used an ultracomplex shRNA screen in the J-lat clone 5A8 also did not identify SETDB1 as being involved in the maintenance of silencing(Besnard *et al.*, 2016).

4.2 Principles of shRNA knockdown of HUSH components

To study the effect of the HUSH complex on HIV gene silencing and latency the HUSH components were knocked down by vectors expressing shRNAs. The vector plasmids used in this study were a kind gift of the Lehner laboratory and express the same HUSH shRNAs used in their paper (Tchasovnikarova *et al.*, 2015).

An shRNA (short hairpin RNA) basically consists of two short (19-22nt) reverse complementary RNA sequences separated by a 4-11nt loop. After transcription the RNA folds such that the reverse complementary sequences pair and the whole structure forms a short hairpin. The structure is then exported to the cytoplasm where the cellular enzyme Dicer removes the loop leaving a short RNA duplex with 2nt overhangs. This dimer is then loaded into the RNA induced silencing complex (RISC) that unwinds the two strands and degrades the coding stand leaving the guide strand intact. The sequence chosen for the shRNA gives the specificity for the target gene. When the guide strand that is captured by RISC binds its target mRNA with perfect complementarity the mRNA is degraded by RISC. If the binding is imperfect, RISC prevents translation of the mRNA product. Both mechanisms result in a reduction in expression of the gene product(Reviewed in Moore et al. 2010).

In contrast to siRNAs, which are transfected into the cytoplasm where they directly interact with Dicer, shRNAs must be transcribed in the target cell. This has the advantage of permitting a more durable knockdown and also of enabling their delivery along with a transduction marker or other selectable protein. Transcription may be achieved by transient transfection with a plasmid expressing the shRNA of choice or by delivering the shRNA for integration into the cellular genome using a lentiviral vector.

Figure 4.1 shows a schematic of the arrangement of the vectors used in these experiments, these encode the shRNA and a transduction tag. The inclusion of a lentiviral packaging signal allows packaging of 'genomic' RNAs generated by transcription from the 5' LTR by HIV Gag. In order to generate lentiviral particles from these plasmids they were co transfected into 293T cells with the packaging plasmids required for a lentiviral system. These were pCMVdeltaR8.91 to provide the HIV derived proteins Gag, Pol, Tat and Rev and pCMV VSVG to provide Vesicular Stomatitis Virus envelope glycoprotein which gives the viral particles a broad cellular tropism by binding to the low density lipoprotein receptor(Finkelshtein *et al.*, 2013).

The SFFV promoter driven SBP- ΔLNGFR codes for a biotinylated version of the truncated Nerve Growth Factor Receptor. ΔLNGFR locates to the cell surface where the biotin permits cell sorting by binding to streptavidin coated magnetic beads (Matheson, Peden and Lehner, 2014). The LNGFR also serves as a transduction marker by allowing the cells to be stained with fluorescent conjugated anti-LNGFR antibodies and evaluated by flow cytometry.

The shRNA sequences used in this study are shown in table 4.1.

Figure 4.1 – Genomic arrangement of the shRNA vector. The packaging signal ψ enables HIV Gag mediated packaging of genomic RNAs driven by the 5' LTR. The Rev response element (RRE) permits Rev binding to RNA and export from the nucleus as well as enhanced packaging. The central polypurine tract (cPPT) is thought to improve nuclear translocation of the reverse transcribed vector during transduction. U6 is a RNA pol III promoter which drives expression of the shRNA. Expression of the transduction tag SBP-ΔLNGFR is driven by a spleen focus forming virus (SFFV) promoter. The woodchuck hepatitis virus posttranscriptional response element (WPRE) forms a structural motif which enhances translation of vector gene products

Target	Sequence
shControl	GATCCACTACCGTTGTTATAGGTGTTCAAGAGACACCTATAACAACGGTAGTTTTTTTG
shMPP8	GATCCAAGAAGACCCCGAGAAAGGTTCAAGAGACCTTTCTCGGGGTCTTCTTTTTTTTG
shPeriphilin	GATCCAGCTAACCACTCGCTCTAATTCAAGAGATTAGAGCGAGTGGTTAGCTTTTTTTG
shTASOR	GATCCGAGGAAGCTTGAGGATCTATTCAAGAGATAGATCCTCAAGCTTCCTCTTTTTTG
shSETDB1	GATCCACCCGAGGCTTTGCTCTTATTCAAGAGATAAGAGCAAAGCCTCGGGTTTTTTTG

Table 4.1 – shRNA sequences used in this study. Each sequence encodes a hairpin motif and contains sequences specific for the gene of interest. The control shRNA is designed not to bind RNA from any cellular genes

4.3 Transduction of cells with shRNA expressing vector

4.3.1 Preliminary experiment - sorting cells by magnetic beads

As described above the vector expressing the shRNAs targeting the HUSH components also express a biotinylated ΔLNGFR. Streptavidin binds biotin tightly but non-covalently. Figure 4.2 shows data obtained by transducing Jurkat cells with the vector expressing a control shRNA. The histogram shows two peaks, an untransduced population and a ΔLNGFR expressing population. The cells were stained with anti-LNGFR APC antibody. The cells were sorted using the magnetic bead system. The magnetic streptavidin coated beads bind the biotin on the surface of transduced cells and enable them to be pulled out by a magnet. The second and third panels show the staining pattern of LNGFR in the positive and negative fractions and the final panel shows how the two populations when overlaid recapitulate the histogram shape of the original mixed population.

4.3.2 Preliminary experiment – Evaluating the shRNA knockdown

Jurkat cells were transduced with the shRNA expressing vectors. 72 hours after transduction the cells were sorted using the magnetic system described above. One million cells were lysed in RIPA buffer. 20ug of lysates quantitated using the Peirce BCA protein assay were loaded into a PAGE gel and analysed by western blot as shown in figure 4.3. The antibodies to MPP8 and TASOR worked satisfactorily and show a knockdown of these proteins at the 72 hour time point. Unfortunately western blotting of Periphilin and SetDB1 proved challenging due to high background signal. Strategies that were attempted to improve detection included trialling a range of concentrations of the primary antibody, using an antibody from a different supplier and changing the duration of blocking. It was not possible to demonstrate the knockdown of these proteins by western blot. However as it had been demonstrated that at least two of the shRNAs were knocking down their target proteins and the other shRNA targets had already been validated by the Lehner group in their 2015 paper it was decided to continue with further experiments.

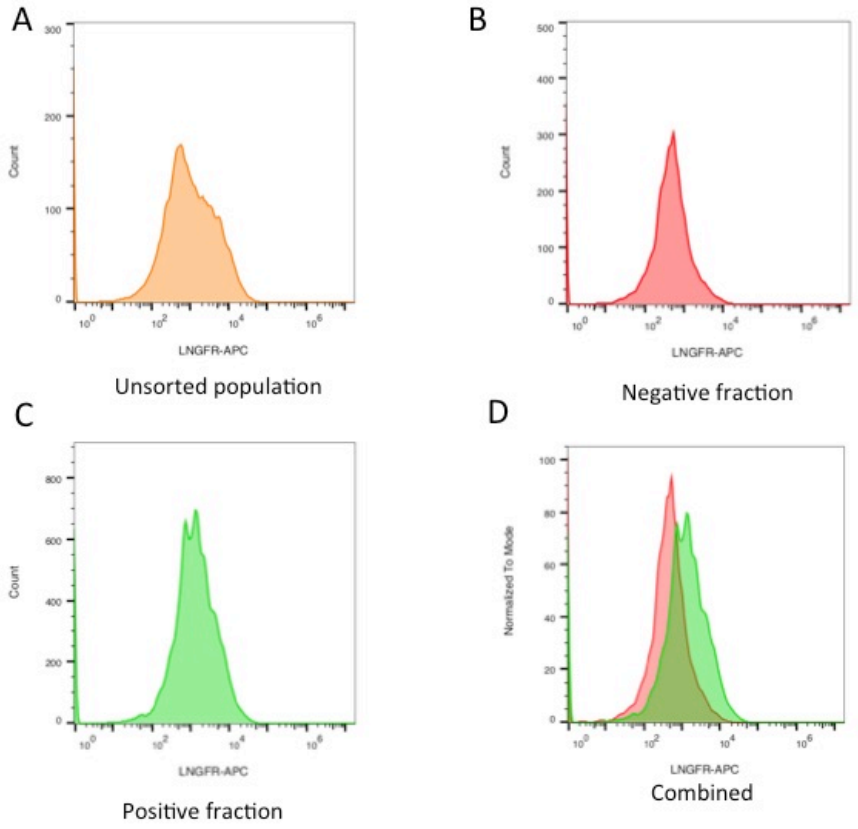


Figure 4.2 – Isolation of transduced cells using magnetic beads. **A** Histogram of distribution of LNGFR staining in a population of cells in which around 50% of cells are transduced. **B** Staining of the 'negative' population after magnetic bead removal of LNGFR expressing cells. **C** LNGFR expression on the positive fraction removed by magnetic beads showing higher levels of expression than **B.** **D** Overlay of the negative and positive fractions showing reconstruction of a similar population structure to **A.**

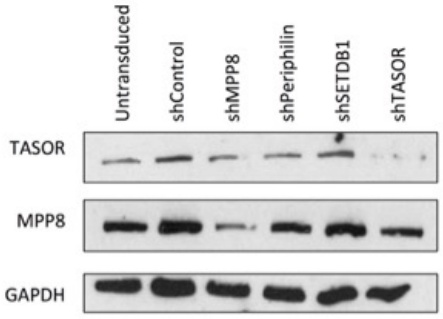


Figure 4.3 – Knockdown of TASOR and MPP8 in shRNA transduced cells. Jurkat cells were transduced with shRNA vectors. After 72 hours transduced cells were isolated using magnetic bead separation. Cells were lysed and proteins analysed by western blot showing reduction in expression of TASOR and MPP8 in the cells transduced with the appropriate shRNA

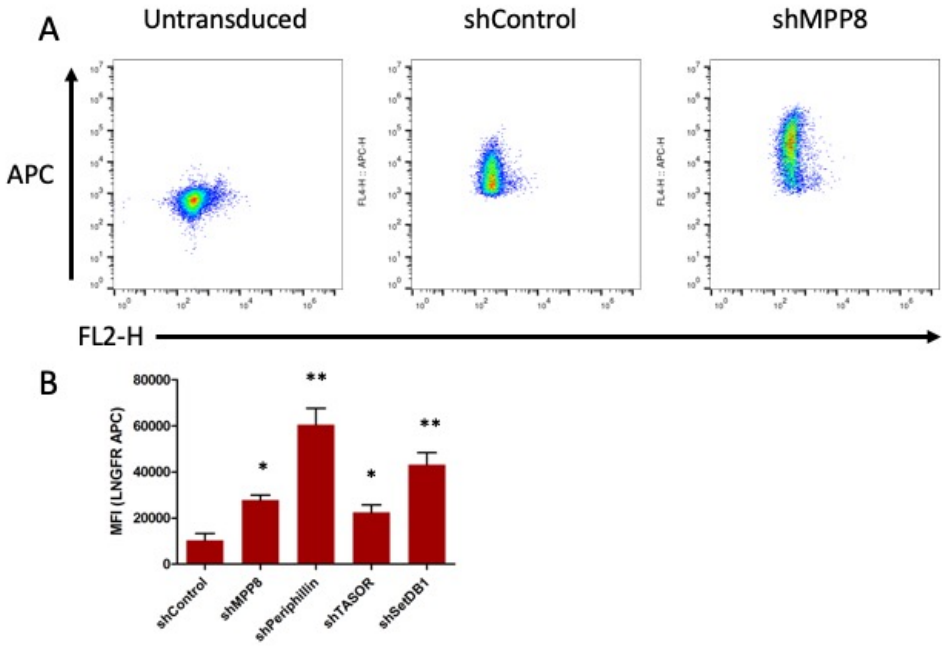


Figure 4.4 – **A** Representative flow plots showing difference in expression of the SFFV driven ΔLNGFR transduction tag in cells transduced with shMPP8 compared to the control shRNA. **B** Higher expression of ΔLNGFR is seen in the HUSH knockdowns. Relative expression of ΔLNGFR measured by mean fluorescence intensity. A greater level of expression was seen in all HUSH shRNAs compared to control. (* $p<0.05$, ** $p<0.01$ by two tailed t test compared to shControl)

4.3.3 HUSH components regulate expression of the SFFV promotor

The amount of ΔLNGFR expressed by the lentiviral vectors was observed to vary depending on the shRNA expressed. Figure 4.4A shows the pattern of staining of shControl and shPeriphilin transduced Jurkat cells, the vertical axis shows the detection of fluorescence from APC conjugated to an anti-LNGFR antibody the horizontal axis shows the signal from the 580/40 filter of the flow cytometer where minimal signal from other fluorophores should be present. Both groups of cells are transduced to a high efficiency (>80% expressing LNGFR) compared to an untransduced control, it is clear that the shPeriphilin transduced cells express a higher level of ΔLNGFR and this can be measured by the mean fluorescence intensity of the cells in the APC channel. This effect was observed for all the HUSH shRNAs as shown in figure 4.4B. These experiments were conducted with identical concentrations of the staining antibody for each condition. One consideration when interpreting these results is that the high MOI used to transduce the cells means that some will have been transduced by multiple proviruses and therefore may produce shRNA at a higher rate than singly transduced cells.

4.4 The effect of HUSH knockdown in J-lat cells

4.4.1 HUSH knockdown in J-lat cells harbouring an LTR-*tat*-IRES-GFP provirus

As noted in the introduction to this chapter the Lehner group examined the effect of HUSH knockdown in four clones of the J-lat model of HIV latency that harbour LTR-*tat*-IRES-GFP proviruses. Two of the clones studied were examined here to see whether the same effects would be observed - the A2 clone and the H2 clone. J-lat cells were transduced with the HUSH shRNA expressing vectors. After 72 hours vector transduced cells were identified by flow cytometry by staining for ΔLNGFR using an untransduced population as a negative control (shown in figure 4.5). The percentage of cells expressing GFP was measured as for the experiments in chapter 3.

Figure 4.6 shows the %GFP expression in vector transduced J-lat cells of the A2 and H2 clones. In addition an untransduced unstimulated negative control and

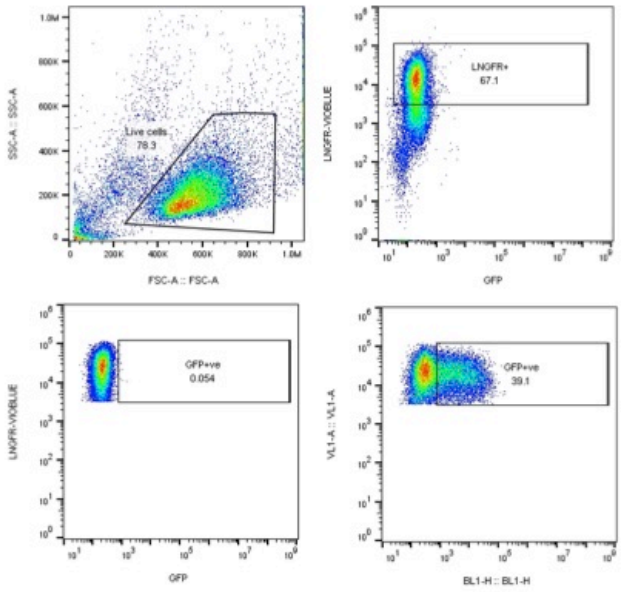


Figure 4.5 – Gating strategy used to identify shRNA transduced cells by flow cytometry. ΔLNGFR expressing cells were identified by staining with Vioblue conjugated anti-LNGFR antibody. Gating on LNGFR positive cells selects for transduced cells. By using a population of non GFP expressing jurkat cells as a negative population a gate identifying GFP positive cells was set

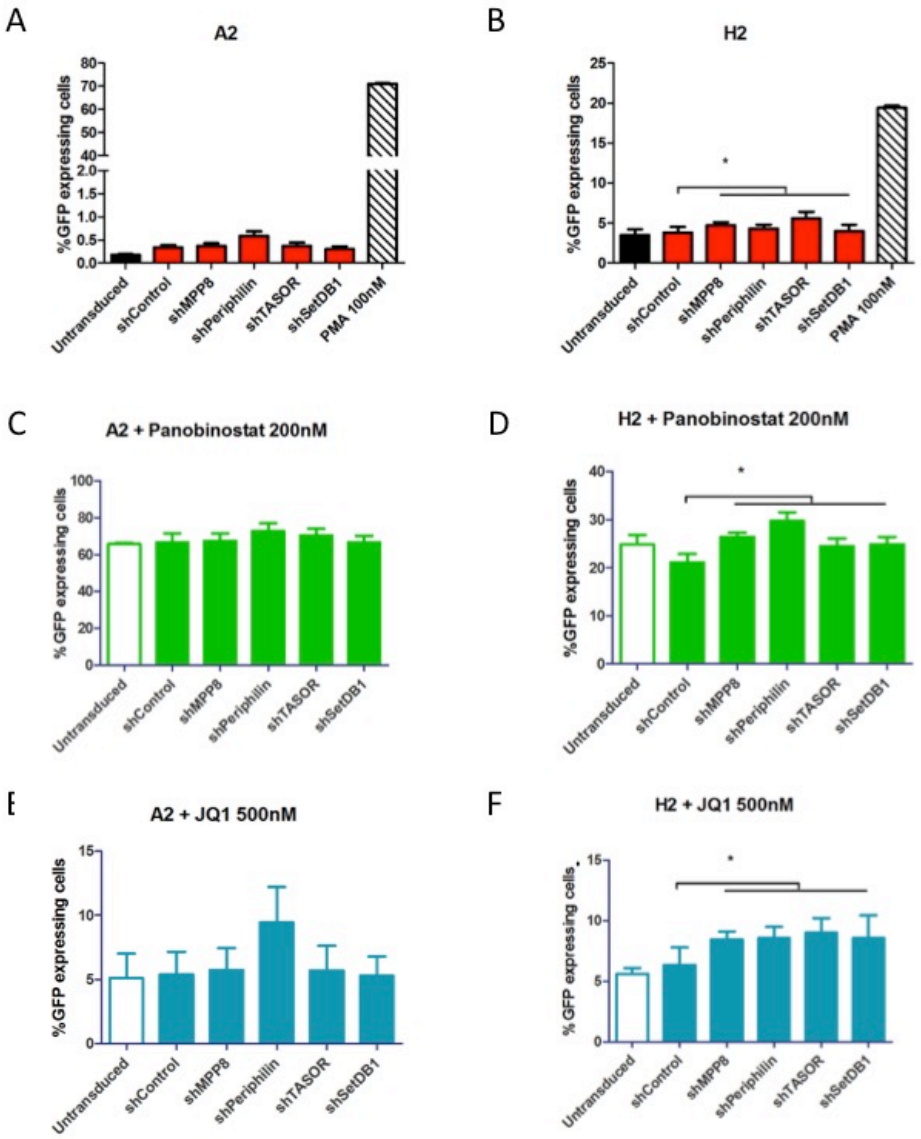


Figure 4.6 – HUSH knockdown increases GFP expression in the J-lat H2 clone. J-lat cells were transduced with HUSH shRNAs and either left unstimulated (**A, B**) stimulated with additional panobinostat 200nM (**C,D**) or stimulated with additional JQ1 500nM (**E,F**). No effect was seen in clone A2. A very modest effect was seen in clone H2 (p=0.38, compared to shControl) this effect was slightly larger in the presence of panobinostat (p=0.02) and JQ1 (p=0.014). Data represent mean and SEM for three experiments.

an untransduced PMA stimulated positive control were included in each experiment. The graphs are representative of the mean and standard error of three independent experiments. Consistent with previously reported findings the A2 clone did not show any activation compared to baseline. The H2 clone showed a very small increase in the number of GFP expressing cells with the shRNA knockdowns compared to the control shRNA ($p=0.038$, two tailed t test). However it should be noted that this effect was extremely small compared to the effect of PMA stimulation. The percentage induced was also smaller than that reported by the Lehner group although they had a baseline expression of approximately 10% in their unstimulated control compared to a baseline of 3.7% in the experiments shown here although a similar fold change in GFP signal was observed here (6.3% compared to 3.7% vs 20% compared to 10%). It is likely that this difference was either due to the method of handling the cells or to the serum used in the culture medium; either way it suggests that additional stimulation can enhance the effect of the HUSH knockdowns.

4.4.2 Additional stimulation with HUSH knockdown

To examine whether HUSH knockdown can enhance the effect of latency reversing agents J-lat cells were transduced with the shRNA vectors as before. After 72 hours the latency reversing agents Panobinostat and JQ1 were added at 200nM and 500nM respectively. The % GFP of the vector transduced cells was determined by flow cytometry as described above. Figure 4.6c and d show the effect of panobinostat in the HUSH transduced J-lats. Once again no effect was observed in the A2 clone. There was a clear increase in panobinostat induced GFP expression in the HUSH knockdowns in the H2 clone when compared to the control shRNA ($p=0.02$). This effect was observed only in comparison to the control shRNA; there was no significant increase over the untransduced control. A similar result was observed with JQ1 (Figure 4.6e and 4.6f), a significant increase in % GFP expressing cells was seen in the HUSH shRNA transduced H2 clone compared to the control shRNA ($p=0.014$).

4.4.3 Effect of HUSH knockdown in J-lat clones with a full provirus

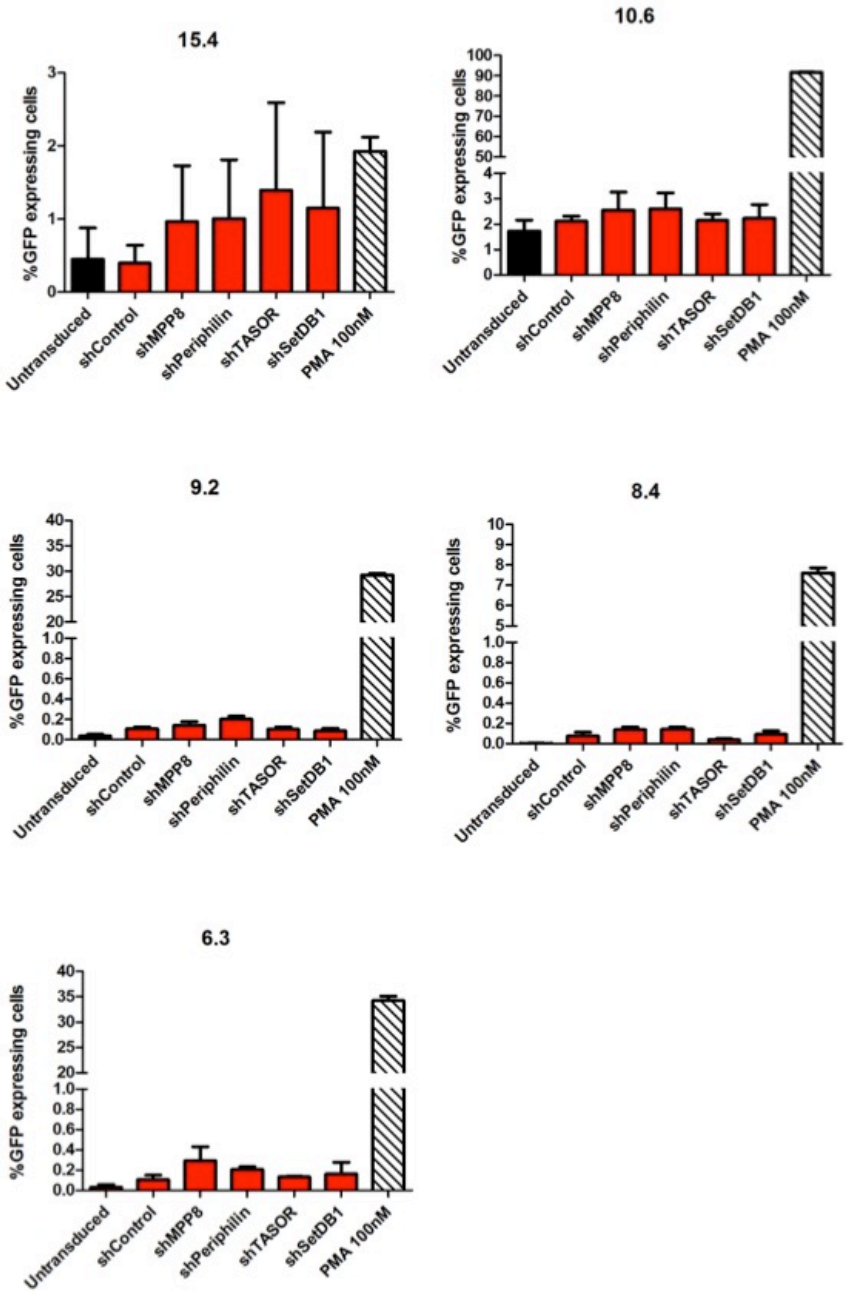


Figure 4.7 – Effect of HUSH shRNAs in full length J-lat clones. No effect was seen in baseline GFP expression in clones transduced with HUSH shRNAs. PMA was included as a positive control. Bars show mean and standard error of three experiments

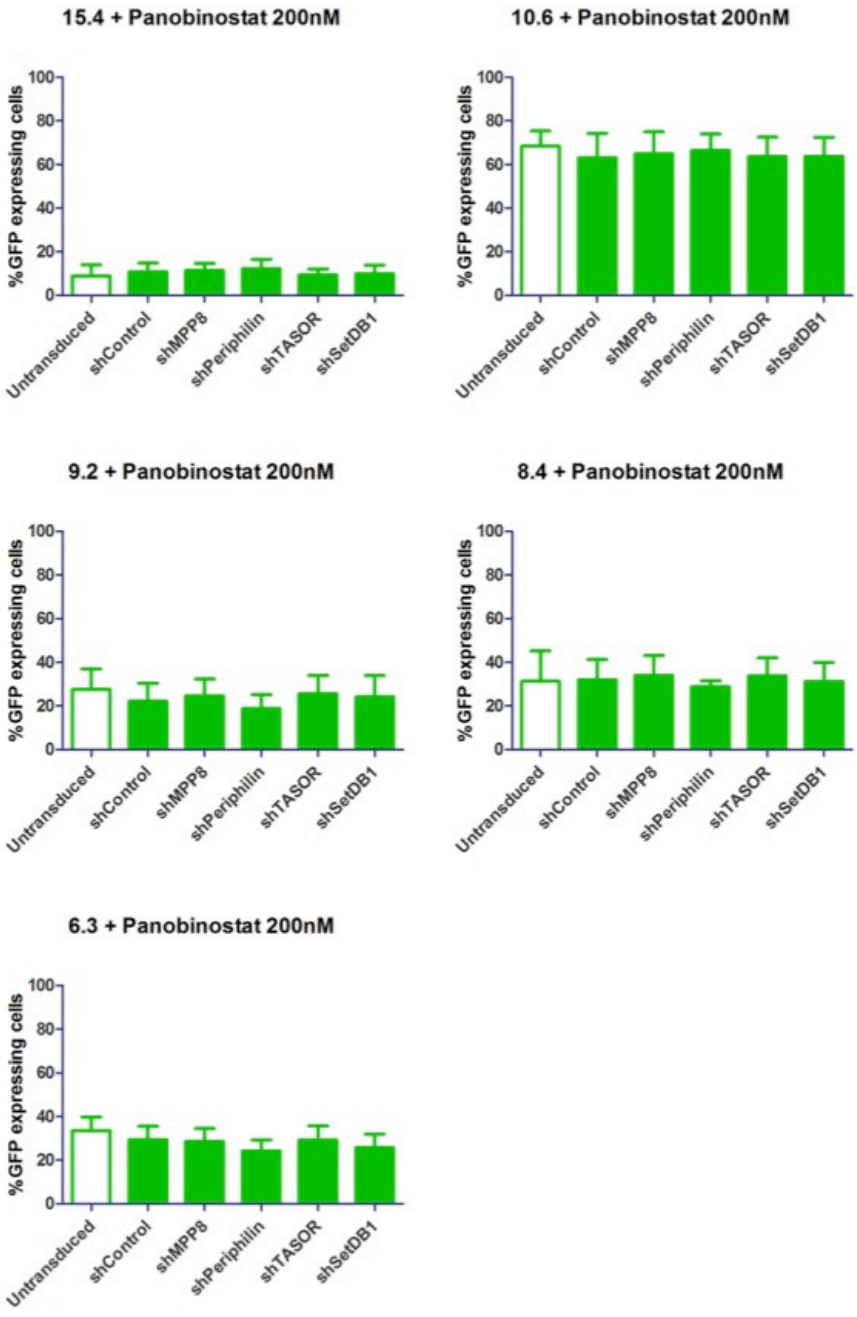


Figure 4.8 – Effect of HUSH shRNAs in full length J-lat clones with additional stimulation with panobinostat 200nM. The untransduced bars show the effect of panobinostat 200nM alone. No significant increase in GFP signal in the HUSH knockdowns in combination with 200nM panobinostat was observed. Data from three experiments, bars show means plus standard error

As the J–lat clones described above harbour an LTR-*tat*-IRES-GFP provirus only it is possible that any effect of HUSH knockdown is enhanced by the fact that the provirus produces only one viral protein, Tat. This means that it may be possible to set up the positive feedback loop of transcription and generate GFP expression at a lower threshold. Therefore the effect of HUSH knockdown in the J lat clones 15.4, 10.6, 9.2, 8.4 and 6.3 was also assessed. These clones each have a more complete HIV provirus with a non functional envelope protein that expresses GFP in place of Nef. These clones were not studied by the Lehner group. The A2 and H2 clones offer a means to evaluate HIV LTR function in near isolation. In comparison the longer provirus means that LTR function is studied in the context of the complex interplay of accessory proteins and splicing factors that affect HIV gene expression.

In an identical way to the experiments described in 4.4.1 the cells were transduced with the HUSH vectors 72 hours prior to read out. The cells were gated according to the strategy in figure 4.5. Figure 4.7 shows the observed % GFP for each of the five clones. A positive control population stimulated with 100nM PMA was also included in each experiment. In no clone was a significant effect of the HUSH knockdown observed and any effect was extremely small compared to the positive control stimulation, suggesting that in these clones HUSH disruption alone was not sufficient to activate the latent provirus. Although clone 15.4 appears to show an effect the data was highly variable, hence wide error bars, and was not statistically significant.

It may be possible that removal of HUSH imposed restriction alone does not trigger transcription from a silenced HIV LTR so the HUSH knockdowns in the full length J-lat clones were also treated with panobinostat 200nM to provide additional stimulation. Figure 4.8 shows the results. No significant effect of the knockdown was observed in combination with panobinostat. As it is possible that the response seen in untransduced J-lats exposed to 200nM panobinostat represents the maximal stimulation (see chapter 3) the experiment was repeated with a reduced dose of panobinostat (20nM) which is less than the EC50 dose for each clone (Chapter 3). Once again no additive effect from the knockdowns was observed (Figure 4.9).

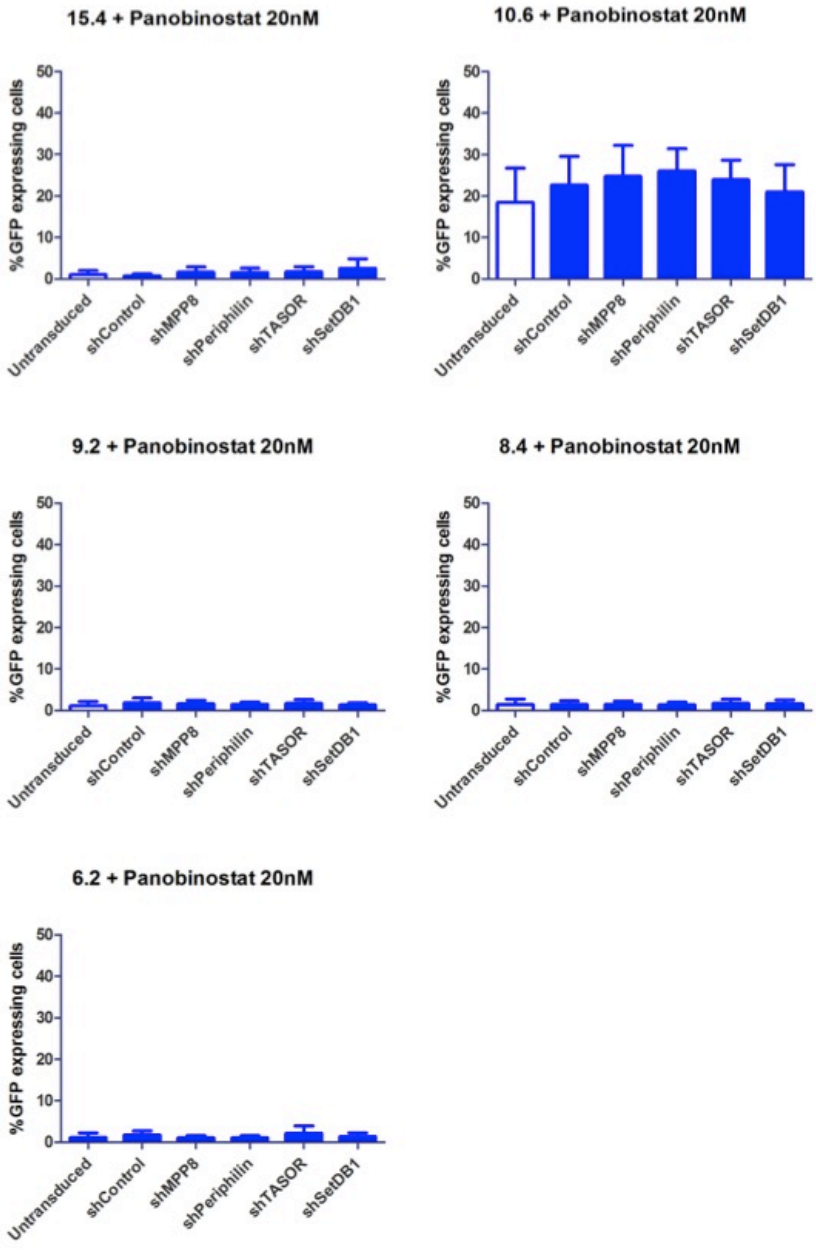


Figure 4.9 – Effect of HUSH shRNAs in full length J-lat clones with additional stimulation with panobinostat 20nM. The untransduced bars show the effect of panobinostat 20nM alone. A reduced dose of panobinostat was used compared to the previous experiment. No effect was observed in any of the clones. Data from three experiments, bars show means plus standard error.

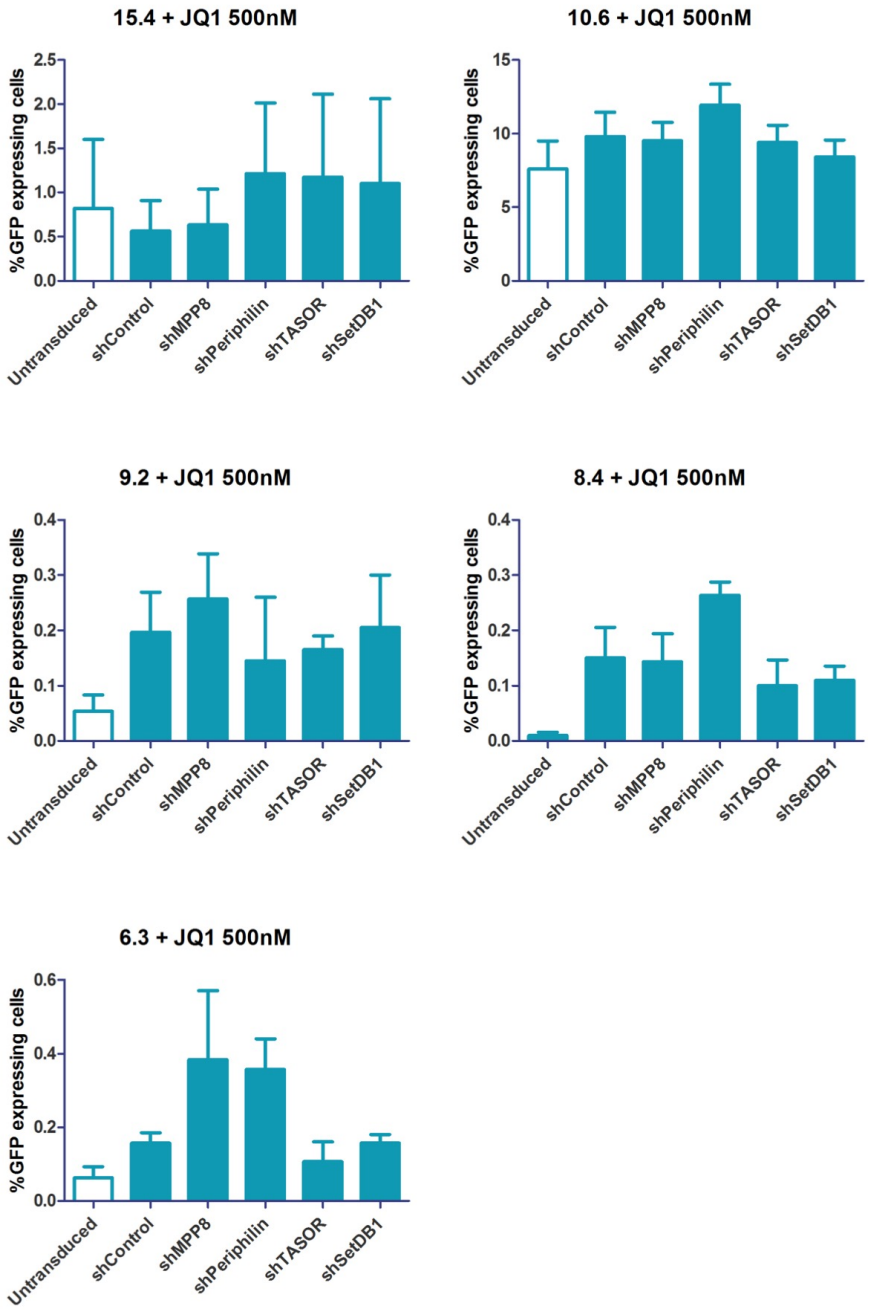


Figure 4.10 – Effect of HUSH shRNAs in full length J-lat clones with additional stimulation with JQ1 500nM. The untransduced bars show the effect of JQ1 alone. No significant effect of HUSH knockdown in combination with JQ1 was observed. Bars show means plus standard error of data from three experiments.

As it is possible that the combination of two mechanisms having a common final pathway of histone modification do not have an additive effect, the same experiment was carried out with 500nM JQ1 as the stimulus. As this provides more PTEF-b to the LTR to enable transcription it would be reasonable to hypothesise an additive effect as removing the block imposed by HUSH may enable the PTEF-b released by JQ1 to activate a greater number of LTRs. The experiment was carried out in the same was as those above. As shows in figure 4.10 no significant effect of the combination of HUSH knockdown and JQ1 simulation was observed.

4.5 The effect of HUSH depletion on expression of incoming virus

As described in the introduction SETDB1 has previously been identified as a possible 'restriction factor' for HIV in an siRNA screen. In this system the knockdown was performed by siRNA transfection prior to infection with a single round HIV vector. To examine whether this effect was limited to SETDB1 or might involve other HUSH partners a similar experiment was carried out. 293T cells were plated into 6 well plates to 80% confluence. Each well was transfected with 4ug of plasmid expressing one of the HUSH shRNAs. After 48 hours the cells were infected with a single round HIV vector NL4.3deltaENV.GFP pseudotyped with VSVG. Each well was infected with the same quantity of virus by dividing the viral stock equally between wells. 24 hours after infection the cells were suspended in enzyme free dissociation buffer to enable the %GFP expressing cells to be counted. Consistent with the findings of Liu *et al* there was a significant increase in the proportion of cells expressing GFP in the SETDB1 knockdown compared to the control shRNA(2.3 fold p=0.008)(figure 4.11). A smaller but still significant effect was seen with the HUSH shRNAs (p=0.01-0.05). This suggests that the effect reported by Liu *et al* could be partially mediated by HUSH recruitment of SETDB1.

To investigate this further an analogous experiment was carried out with Jurkat cells. Jurkat cells were transduced with the HUSH shRNA vectors. After 72 hours the cells were sorted using the magnetic bead system described in 4.3.1 to obtain

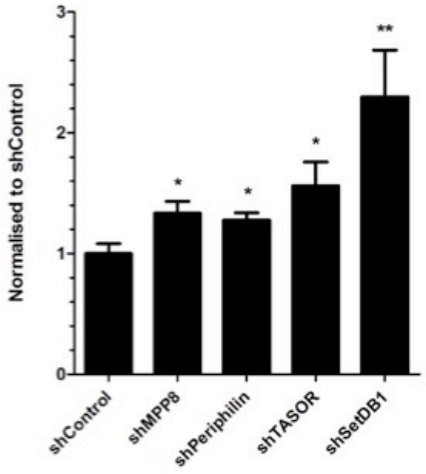


Figure 4.11 – Effect of HUSH shRNAs on incoming virus in 293T cells. Cells were transfected with HUSH shRNA expressing plasmids in 6 well plates. Equal volumes of NL4.3 GFP virus stock were distributed into each well. % of cells expressing GFP was measured. Data was normalised to the GFP expression in the control shRNA group. There was a 2.3 fold increase in GFP expression in the shSETDB1 cells compared to shControl (p=0.008, two tailed t test (**)). A smaller but still statistically significant effect was seen in the HUSH shRNA groups (* p=0.01-0.05). Bars represent mean and standard error of 4 experiments

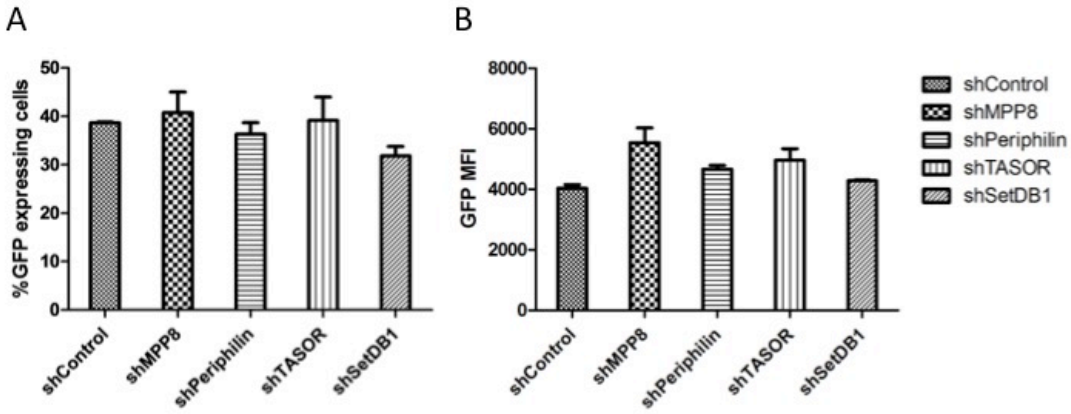


Figure 4.12 – Effect of HUSH shRNAs on incoming virus in jurkat cells. Cells were transduced with HUSH shRNA vectors. After sorting the cells using magnetic beads the cells were transduced with equal volumes of NL4.3 GFP virus. No difference was seen in the proportion of cells which expressed GFP (**A**) or in the mean fluorescence intensity of the GFP signal in infected cells (**B**). Data from three experiments

populations of transduced cells. The cells were counted and infected with VSVG pseudotyped NL4.3deltaENV.GFP at identical MOIs. After 48 hours the cells were analysed for GFP expression. The results are shown in figure 4.12, In contrast to the findings in 293T cells there was no difference in the percentage of GFP expressing cells in the HUSH shRNA transduced cells compared to the control shRNA. There was also no difference in the mean fluorescence intensity suggesting that there was no difference in the rate of GFP transcript production in the HUSH shRNA transduced cells. It should be noted that protein levels in 293T cells were not evaluated here, it is therefore not possible to exclude a difference in the efficiency of the knockdown between 293T cells and Jurkat cells.

To explore the effect of HUSH knockdowns on LTR silencing a VSVG pseudotyped dual fluorescent virus (named RGH) was employed (Dahabieh *et al.*, 2013). This virus expresses LTR driven GFP in the Gag open reading frame and a CMV driven mCherry in place of Nef (figure 4.13a). Using this model cells with integrated virus express mCherry (red) and those with active LTR transcription also express GFP (green) therefore the proportion of single red and red/green double positives gives an estimate of the proportion of integrated proviruses with silenced LTRs. Figure 4.13b shows an example plot of Jurkat cells infected with the RGH virus.

In a similar system to that described above HUSH transduced Jurkat cells were infected with the RGH virus 72 hours after transduction. Figure 4.13c shows example flow plots from this series of experiments. No difference in the proportion of red and red/green cells was seen between the control shRNA and the HUSH shRNAs. Unexpectedly a population of GFP single positive cells was observed which was not seen in non shRNA transduced cells (figure 4.13b). This population was observed to become more brightly GFP positive (by mean fluorescence intensity) in the HUSH shRNA transduced cells (figure 4.13d). In their paper describing their RGH vector the group only found GFP single positives in the first 24 hours and they attributed these to GFP-Gag fusion products produced in the vector preparation. They deduced this by

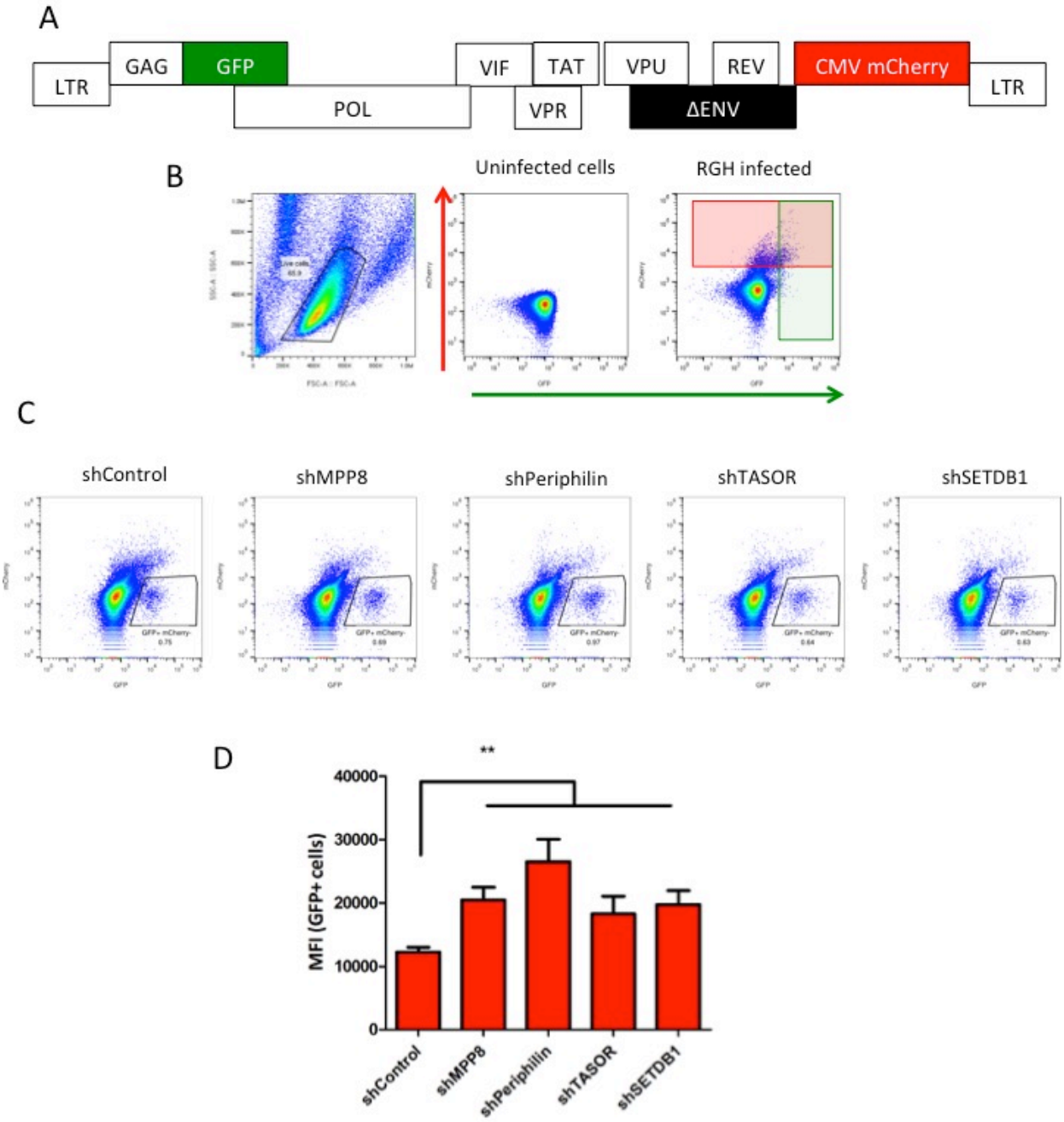


Figure 4.13 – Effect of HUSH shRNAs on a dual fluorescent virus in jurkat cells. Cells were transduced with HUSH shRNA vectors. **A** Arrangement of dual fluorescent vector RGH. GFP is expressed in Gag, mCherry is expressed in place of Nef and is driven by its own CMV promoter. The virus has an envelope deletion to prevent multiple rounds of infection. **B** Gating diagram showing cells transduced with RGH expressing mCherry and GFP. **C** Example flow plots showing the observed GFP positive mCherry negative population in cells transduced with HUSH vectors **D** Mean fluorescence intensity of the GFP single positive population is increased in cells transduced with HUSH knockdown (p=0.008, two tailed t test). Data from three experiments, bars show mean and standard error

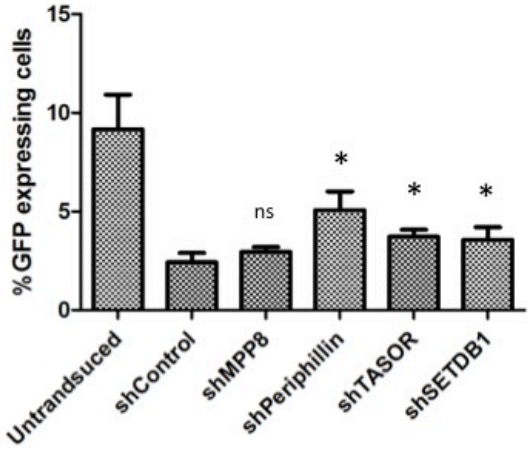


Figure 4.14 – HUSH knockdown in infected cells increases GFP expression. Jurkat cells were infected with NL4.3 GFP. After 48 hours the cells were transduced with HUSH shRNA vectors. A small but statistically significant increase in GFP signal was seen in cells transduced with shRNAs to Periphilin, TASOR and SETDB1 compared to shControl (* = $p<0.05$, two tailed t test). The small increase seen in shMPP8 was not statistically significant. GFP expression in all shRNA transduced cells was lower than the untransduced control.

demonstrating that the GFP signal persists despite inhibition of integration. They did not specifically address the possibility that GFP production could come from unintegrated HIV DNA. The observation here of an increase in GFP signal strength suggests an increase in production of GFP in the absence of mCherry production in the HUSH shRNA transduced cells. The alternative explanation is that the GFP expressing cells received more virions and therefore more GFP-Gag fusion protein than uninfected cells although this seems unlikely as this would also predict a greater number of GFP positive cells rather than the small population seen here.

4.6 Effect of HUSH shRNA knockdown on cells transduced with GFP virus

It is possible that HUSH acts early in the silencing process but is not required for long term maintenance of silencing. To examine whether silenced proviruses from a recent infection could be reactivated by disruption of HUSH, Jurkat cells were infected with NL4.3.deltaENV.GFP (10ng p24 in 10^6 cells). After 48 hours the cells were transduced with vectors expressing the HUSH shRNAs. After a further 48 hours the cells were examined by flow cytometry. As before shRNA transduced cells were identified by staining for LNGFR and the percentage of GFP expressing cells determined. Figure 4.14 shows the GFP signal observed. Interestingly all the HUSH transduced cells showed a lower proportion of GFP expressing cells than the untransduced control. One explanation for this is that the lentiviral vector preferentially transduced the uninfected cells; resistance to superinfection is a phenomenon that has been described in both HIV (reviewed in Nethe et al. 2005) and in lentiviral vector systems (Liao *et al.*, 2017)but is not fully understood. An alternative explanation might be that transduction with an shRNA expressing vector supresses HIV transcription.

Nonetheless there was an increase in the proportion of GFP expressing cells in the HUSH shRNA knockdowns compared to the shRNA control. It is possible that this signal comes from integrated proviruses however it is important to note that it was not possible to include antiretrovirals in the culture medium as this would prevent transduction by the lentiviral vector. Therefore it is not possible to

exclude an effect from the shRNA vectors making the cells more permissive to infection by any residual GFP virus in the supernatant or from cells harbouring viral genomic material that is either incompletely reverse transcribed or has not been integrated. One approach to reconcile this would be to use a vector with mutations in RT and integrase which confer resistance to efavirenz and raltegravir

4.7 Effect of incoming virus on HUSH protein levels

One possibility to explain why no effect on HIV expression levels was seen in the HUSH shRNA transduced Jurkat cells (figure 4.12) is that infection by HIV modulates the levels of HUSH. This might be expected if HUSH functions as a significant restriction factor *in vivo* as the pressure exerted by this would promote viral mechanisms to prevent repression by HUSH. Viral evasion of host restriction factors has been described previously for example the host factor APOBEC which introduces hypermutation in incoming viruses is degraded by the viral protein Vif (Stopak *et al.*, 2003; Binka *et al.*, 2012). To explore whether this may be happening in the system described here Jurkat cells were infected with NL4.3.deltaENV.GFP pseudotyped with VSV glycoprotein. An uninfected control as well as cells infected in the presence of 15nM efavirenz or 100nM raltegravir were also included. Protein from cells was harvested 48 hours after infection. The levels of MPP8 and TASOR were analysed by western blot. 20ug of protein, normalised by pierce BCA assay, was loaded per sample per lane and therefore equal volumes of lysate were loaded for both the MPP8 and TASOR blots. The blots were divided and also probed for GAPDH as a loading control. Protein levels were determined by densitometry of scanned images of the x ray film. Protein levels were normalised to the density of the GAPDH band. Figure 4.15 shows the results of 4 experiments, there was a significant reduction in the levels of both TASOR (0.3 vs 1, p=0.006) and MPP8 (0.48 vs 1, p=0.01) which was abrogated by the inclusion of antiretrovirals.

Previous studies of protein levels after infection with HIV have not identified changes in the levels of these proteins (Greenwood *et al.*, 2016) therefore it was

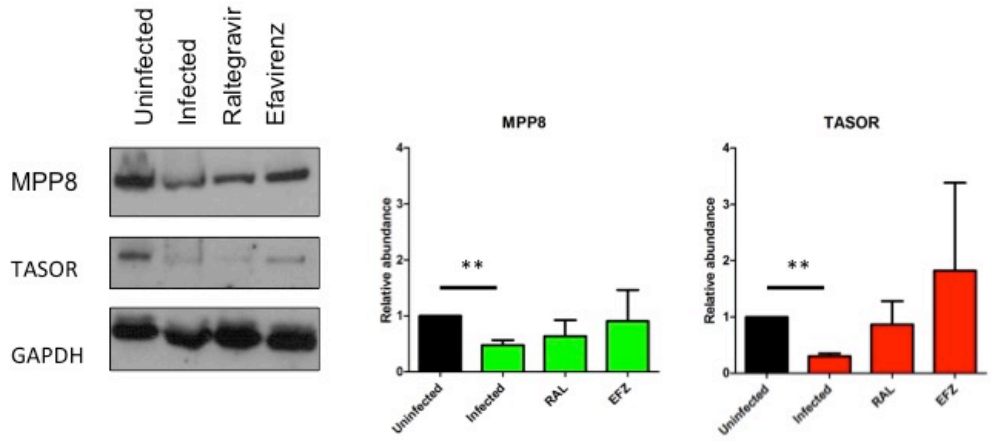


Figure 4.15 – Expression of MPP8 and TASOR analysed by western blot appears reduced in cells infected with NL4.3 GFP. Jurkat cells were infected at high MOI with NL4.3 GFP. Cells were also infected in the presence of raltegravir 100nM and efavirenz 15nM. Cells were lysed 48 hours after infection. 20ug of protein lysate was analysed by western blot. A significant reduction in both TASOR and MPP8 was seen (p=0.01 and p=0.006 respectively, two tailed t test). Results represent 4 experiments

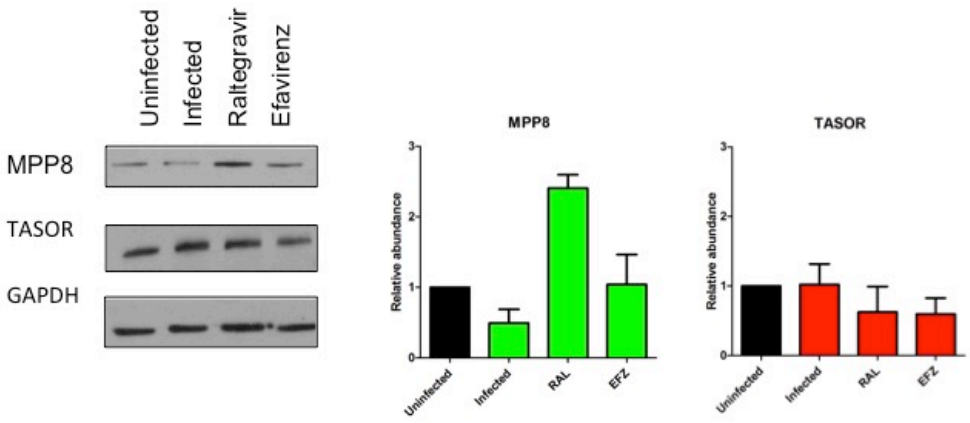


Figure 4.16 – Expression of MPP8 and TASOR is not reduced in cells infected with replicating virus. Jurkat cells were infected with replicating $HIV_{NL4.3.}$ cells were also infected in the presence of either raltegravir 100nM or efavirenz 15nM. Cells were lysed after 48 hours and protein concentrations normalised. 20ug of protein was loaded in each lane. In contrast to the previous experiment, no significant effect on protein levels was seen.

possible that the changes observed above were due to the pseudotyped vector used or to cell death induced by infection. To explore this possibility the experiment was repeated using NL4.3 with a native envelope and at a lower MOI (0.1). Again uninfected cells and cells infected in the presence of raltegravir 100nM and efavirenz 15nM were included, protein was harvested and analysed by western blot using the same conditions outlined above. The results of 3 experiments are shown in figure 4.16. No statistically significant difference was observed. It is possible that the effect observed in figure 4.15 was due to the VSV envelope or due to a higher MOI. Also it is possible that the effect is true however a lower proportion of cells were infected in the replicating virus experiment meaning that the effect was not measurable by western blot. The wide error bars in both experiments also highlight the limitations of the technique used. Detection of protein by western blot and chemiluminesence is prone to problems with the linearity of the detection medium as well as issues with oversaturating a signal.

4.8 Conclusions and discussion

It was not possible to demonstrate a clear role for HUSH mediated recruitment of SETDB1 in HIV latency in these experiments. While the HUSH shRNA knockdowns clearly had some effects, notably on the SFFV promoter driving LNGFR expression and in 293T cells, depletion of HUSH did not affect the reactivation of the five J-lat clones with full length proviruses which were examined here. Possible reasons for this discrepancy are explored below.

A small degree of activation was seen in the H2 J-lat clone when combined with other activating stimulus. This is similar to the results presented by the Lehner lab in their paper. It should be noted however that no signal was seen in the absence of additional stimulation whereas the Lehner lab did report an increase in unstimulated cells. One possible explanation could be found in the high background signal reported by the Lehner lab (around 10% GFP +ve) compared to the experiments reported here. This is likely to be an effect of differences in cell culture medium including the serum used, as such the cells may have experience

additional stimulation from the FCS which provided an increased signal in their experiments.

An effect of HUSH and SETDB1 knockdown similar to that described by the McKnight lab was observed in 293T cells. The observation that knockdown of the HUSH components gives rise to a similar though lesser effect than SETDB1 suggest that they may play a partial role in the SETDB1 effect though modulating its recruitment. The McKnight lab demonstrated a three fold increase in cell associated integrated HIV DNA on knockdown of SETDB1 and postulated that it has a role in restricting integration of incoming proviruses. A similar effect was not seen here in Jurkat cells. They may be due to some intrinsic difference between cell types or due to the difference in shRNA production used (transfection of 293T cells vs transduction of Jurkats). The relative amount of integrated HIV DNA was not explored here although it would have been a useful addition to further confirm the work of the McKnight lab.

The J-lat clones in this study did not respond to HUSH depletion. This is consistent with those siRNA screens performed in J-lats to specifically look for chromatin modelling proteins maintaining HIV latency which did not find SETDB1 as a 'hit'. One possibility for the difference between the full length J-lat clones, which do not respond to HUSH disruption, and the LTR-*tat*-IRES-GFP clones, which do, is differences in integration site. The proviruses in the full length clones are integrated within gene encoding regions(Jordan, Bisgrove and Verdin, 2003) as is the A2 clone; the H2 clone on the other hand has its provirus in a centromeric alphoid repeat(Gallastegui *et al.*, 2011). The centromeric region is known to be rich in H3K9me3 and the Lehner lab demonstrated that this is the signal for HUSH mediated SETDB1 recruitment, it is therefore possible that this integration site is more vulnerable to being shut down by HUSH due to its genomic location.

One further consideration is that the screen in which HUSH was identified was carried out by identifying 'dim' GFP expressors in KBM7 cells and then knocking out proteins and observing for those that turned 'bright'. This system identifies factors causing low expression or partial repression that is not precisely

analogous to HIV latency where it is generally accepted there is no baseline LTR expression. It is therefore possible that HUSH is involved in the early switch off process and the 'dim' population tested corresponds to cells transitioning to silencing. In contrast if HUSH is not involved in the maintenance of silencing it would be expected that depletion of this alone does not reactivate silenced proviruses. Also, as mentioned in the introduction to this chapter SETDB1 mediated methylation of Tat has been described, this is an alternate mechanism by which a 'dim' signal might become 'bright' on depletion of SETDB1.

One intriguing possibility from the results in this chapter is that HUSH and SETDB1 may act at a step prior to integration. This would be consistent with the data from figure 4.12b and with the hypothesis in Liu *et al.* that SETDB1 has a role in preventing integration. The data from the dual fluorescent vector presented here suggest an effect on unintegrated species (figure 4.14) which could either be due to enhanced transcription from a circularised HIV genome or to an increase in HIV DNA. SETDB1 has been shown to act on episomal DNA in the context of the covalently closed circular DNA of hepatitis B virus(Rivière *et al.*, 2015).

Together these may explain why a small signal was observed in cells that had recently been infected with GFP expressing HIV (figure 4.15). The cells could have harboured both integrated proviruses that were transitioning to latency as well as unintegrated HIV DNA. One way in which to evaluate any effect on preintegration HIV DNA would be to study HUSH disruption in the context of a GFP expressing HIV virus with a mutant integrase.

One limitation of the studies here is that it was not possible to confirm a knockdown for all the HUSH proteins by western blot as high background signal in the blots for SETDB1 and Periphilin prevented their detection. It is therefore possible that the knockdown was not sufficiently complete to have an effect. In mitigation it is worth noting that a clear effect was observed on the SFFV promoters in the vector. One way in which the certainty of isolating transduced cells could have been improved would have been to use a puromycin selectable vector. This was not done originally as products released from the dying cells

could have influenced the activation state of other cells in culture and given a false signal. Also a puromycin selectable system was not used as part of the initial plan for this project was to move into studying primary cells and it was thought that the use of an antibody free sorting system would permit isolating transduced cells without changing their activation status.

Chapter 5: The effect of mutations in ESE_{tat} on HIV silencing

5.1 Background

As noted in chapter 1 there is evidence to suggest that some latently infected cells harbour intact proviruses which are not induced upon stimulation (Ho *et al.*, 2013). These viruses do not contain large deletions that would be expected to prevent effective production of new virions. Ho *et al*. cloned a small number of these full length non induced viruses and demonstrated that they did not have a replication defect. The fact that these intact viruses were not reactivated suggests that either the genomic locus into which they were inserted was tightly repressed and therefore not permissive for transcription from the provirus, or that these viruses had a more subtle phenotype which altered silencing behaviour and changed the threshold for reactivation from latency.

Cellular responses that silence proviral transcription by modification of the local histone environment or by altering the availability of critical transcription factors are well studied (summarised in chapter 1). In contrast little is known about virus mediated effects that could alter silencing behaviour. From an evolutionary point of view there seems to be little advantage to HIV in mutating to enhance latency however some authors have suggested advantages during early infection when there are relatively few virions(Rouzine, Weinberger and Weinberger, 2015). In the context of continued antiretroviral suppression, however, those viruses that remain latent could be expected to persist in the latent pool for longer. Examples of viruses that exhibit enhanced silencing behaviour have previously been described. The widely used cellular model of HIV latency, ACH-2, is based on the CEM cell line infected with the laboratory virus HIV_{LAV}(Folks *et al.*, 1989). The provirus in ACH-2 was later shown to have a point mutation in the TAR loop which abrogates binding of Tat and prevents transactivation. Another mutation which has been shown to enhance silencing is H13L Tat (Tyagi, Pearson and Karn, 2010).

This mutant Tat protein is less able to bind PTEF-b; it is capable of some transactivation but is not as efficient as wild type Tat.

Both the examples above affect the Tat-TAR axis that is essential for efficient transcription from the HIV provirus. As it is so key to transcription perturbations in this feedback loop are logical places to look for mutations that enhance silencing. It should be noted that neither of the cases above is found to occur *in vivo.*

Recently an exonic splice enhancer (ESE) responsible for balanced splicing of *tat* mRNA has been identified (Erkelenz *et al.*, 2015) in a region corresponding to HXB2 sequence 5807-5838. This region was located as being downstream of the A3 splice acceptor within the Tat coding region. The authors of the study identified ESE_{tat} using software which would predict splice factor binding sites(Erkelenz *et al.*, 2014). The same software was used to identify a maximally deleterious mutation which would abrogate splice factor binding and disrupt *tat* mRNA splicing. They demonstrated that their mutant produced singly spliced mRNA species but very little *tat* mRNA. This led to a severe replication defect and very limited protein production.

Alternative splicing allows HIV to encode multiple proteins from a single RNA(Ocwieja *et al.*, 2012). It is reliant on the cellular splicing machinery to enable this. HIV mRNA is spliced in the same way as cellular genes; splicing occurs between 5' splice donors and 3' acceptors. In simple splicing the process starts with the binding of the U1 small nuclear RNA (snRNA) to the 5' splice site. The binding of U2 auxiliary factor and U2 snRNP to the branch point and 3' splice site defines the intron to be removed during splicing (Figure 5.1). The binding of the tri-snRNP (U4,U5,U6) catalyses the formation of a loop structure from the intron by cleaving the 5' splice site and joining it to the branch site. The 3' splice site is then cleaved and the exons ligated.

Alternative splicing is more complex as multiple products are formed from a single precursor. Exonic splice enhancers (ESE) and suppressors regulate the

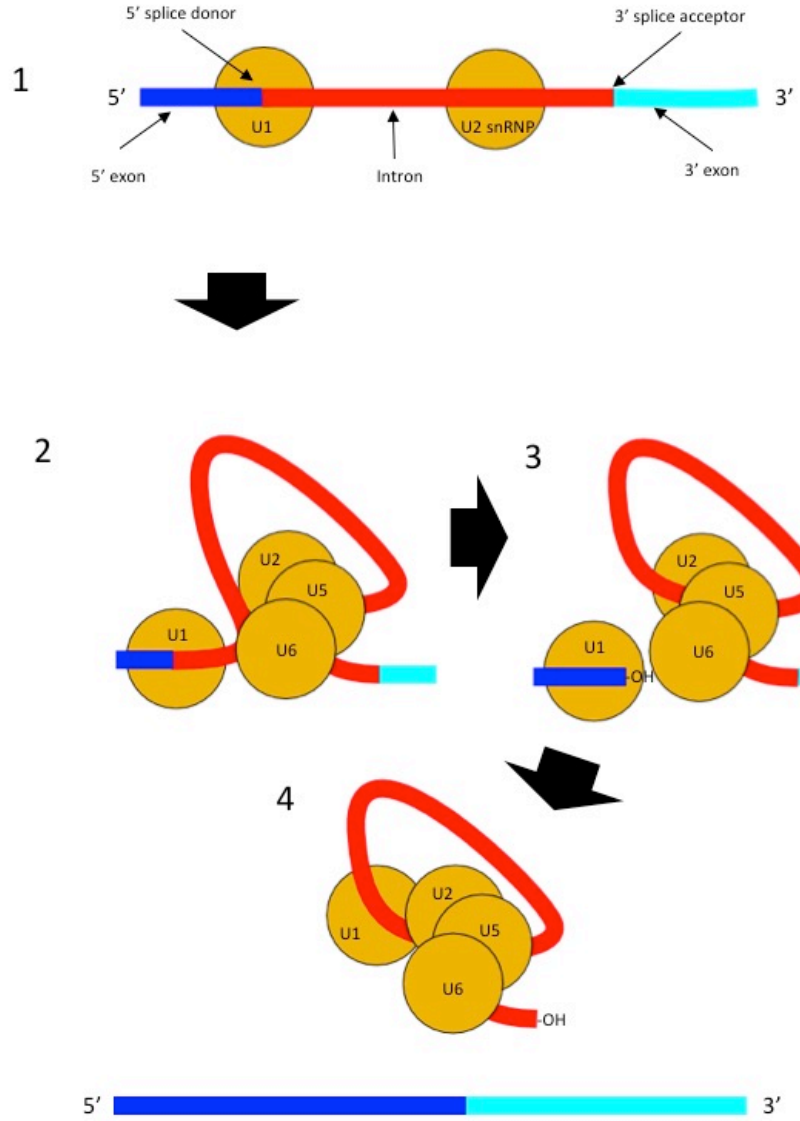


Figure 5.1 – Schematic of splicing of a pair of exons. 1. U1 snRNP binds to the 5' splice site and the U2 snRNP binds to a site upstream of the 3' splice site which will become the branch point. 2. The binding of U4/U5/U6 catalyses the formation of a loop. 3. The branch point attacks the 5' splice site allowing the free 5' end to join the branch point and form a lariat. 4. The 3' end of the 5' exon is ligated to the 3' splice site allowing the intron to be completely excised

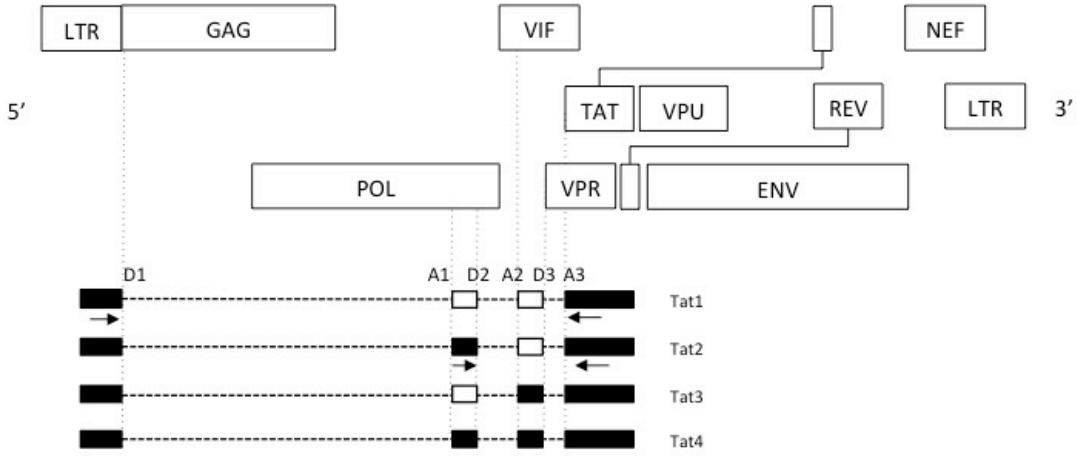


Figure 5.2 – Schematic diagram of the alternate splicing of the major isoforms of HIV *tat* mRNA. The inclusion of two small exons enables the formation of four different isoforms of *tat* mRNA. The arrows show the approximate locations of the primer pairs used to amplify splice junctions unique to Tat1 and Tat2.

use of splice donors and acceptors and therefore regulate exon choice by altering the binding of splicing factors. In general terms the SR proteins recognise specific RNA sequences within ESEs for binding. Once bound they recruit U2AF and the U2 snRNP to the 3' splice acceptor. In the case of ESE_{tat} Erklenz *et al*. showed that the splice enhancing activity was dependent on the binding of Arginine-Serine Rich Splicing Factor 2 (SRSF2).

HIV *tat* mRNA is multiply spliced and can have a number of different isoforms depending on the inclusion of small exons (Figure 5.2). The predominant isoform (*tat* 1) is formed from the joining of the major splice donor D1 to the A3 splice acceptor. All *tat* mRNA species are translated from the same start codon, the length of the protein varies according to the splicing of the first and second coding exons. The functional difference between the various *tat* mRNA isoforms is not known. A recent in depth analysis of splice site usage utilising deep sequencing found a wide range of D1/A3 splicing in a panel of subtype B transmitted/founder viruses compared to an NL4.3 control (Emery *et al.*, 2017) although the authors did not explore the factors leading to altered splice junction use.

5.2 Polymorphisms can be found in the ESEtat region of subtype B sequences

To examine the whether polymorphisms in the ESEtat region occur *in vivo* 2013 subtype B sequences from the Los Alamos national database were examined. Table 5.1 shows the percentage base identity for each position in ESEtat. The majority of positions show greater than 90% conservation of a particular base however there are positions that demonstrate a higher degree of polymorphism. This suggests that evolutionary pressure constrains sequences in the ESEtat region but minor sequence variants exist.

5.3 Predicted binding of splicing factors to mutant ESEtat regions

To determine whether ESEtat regions containing mutant sequences might have altered splicing activity it is possible to predict splice factor binding using the PESX score(Zhang and Chasin, 2004). The score is derived by considering the input

A

Subtype B sequences from LANL (n=2013)

A	1	83	4	2	0	21	79	1	75	12	97	0	17	100	0	98	28	0	87	100	0	90	99	100	0	0	0	100	6	0	0
C	14	8	44	3	99	21	9	90	1	0	0	1	0	0	0	0	1	94	7	0	0	0	0	0	0	0	0	0	0	100	100
G	0	1	1	3	0	47	2	1	15	88	1	97	82	0	99	2	70	1	5	0	100	10	0	0	0	100	100	0	81	0	0
T	85	7	49	91	0	9	2	1	1	0	0	0	0	0	0	0	1	5	0	0	0	0	0	0	100	0	0	0	13	0	0

B

WT	T	A	C	T	C	G	A	C	A	G	A	G	G	A	G	A	G	C	A	A	G	A	A	A	T	G	G	A	G	C	C
M1	T	A	**A**	T	C	**C**	A	C	**G**	**A**	A	G	G	A	G	A	**A**	C	A	A	G	A	A	A	T	G	G	A	G	C	C
M2	T	A	C	T	C	**C**	A	C	A	**A**	A	G	G	A	G	A	G	C	A	A	G	A	A	A	T	G	G	A	**T**	C	C
ERK	T	A	C	T	C	G	A	C	A	**T**	A	G	G	A	**T**	A	G	C	A	A	**A**	A	A	A	T	G	G	A	G	C	C

Table 5.1 – **A** Percentage base identity in ESEtat in 2013 subtype B sequences from the Los Alamos database. Some regions of ESEtat are highly conserved however there are also bases with a high rate of polymorphism. **B** ESEtat regions examined in this study. The top line shows the wild type ESEtat sequence found in $HIV_{NL4.3}$ M1 and M2 represent sequences identified from full length sequences isolated from patients and deposited in the database. Neither M1 nor M2 contain mutations at any of the bases showing 100% conservation. The ERK sequence represents the mutation introduced by Erkelenz et al. to maximally abrogate the function of ESEtat.

sequence as a collection of overlapping octamers. The score for each octamer was derived by Zhang and Chasin using a computational analysis of octamers found to be over represented in non-coding exons compared to pseudo exons. The authors of the study also cloned putative exonic splice enhancers identified by their approach into a mini-gene and demonstrated that they function as enhancers of splicing.

Figure 5.3 shows P scores derived using an online tool which utilises the PESX score to predict splice factor binding (available from http://cubio.biology.columbia.edu/pesx/pesx/). The dashed line in each figure shows the score for the wild type sequence. The horizontal line at p=2.5 represents the cut off proposed by Zhang and Chasin for the identification of putative exonic splice enhancers. It can be seen that the wild type sequence has a peak that is above this line showing that the predictive software confirms the findings of Erkelenz *et al*. The red lines in the figure show the P scores for mutant ESEtat sequences. The ESEtat region from full length subtype B sequences were analysed by hand to identify a sequence predicted to exhibit higher splice factor binding (identified as M1 here) and a sequences predicted to exhibit reduced splice factor binding (M2). The ERK mutant represents the same mutation created by Erkelenz *et al*. to maximally abrogate ESEtat function. The difference in P scores between wild type and mutant was 13.9, -5.3 and -55.8 for the M1, M2 and ERK sequences respectively.

5.4 Effect of mutant ESEtat regions on *tat* splicing

To examine the effect of the ESE sequences identified above they were cloned into the laboratory viruses $HIV_{NL4.3}$ and NL4.3deltaENV.EGFP. The plasmids encoding these viruses are 15kb long which is at the upper limit for site directed mutagenesis therefore a 1.8kb fragment containing the ESEtat region was excised from the plasmids using the enzymes *Eco*RI and *Nhe*I and inserted into the corresponding restriction sites in the cloning plasmid pBR322. Once in the pBR322 plasmid backbone the sequences were altered using site directed

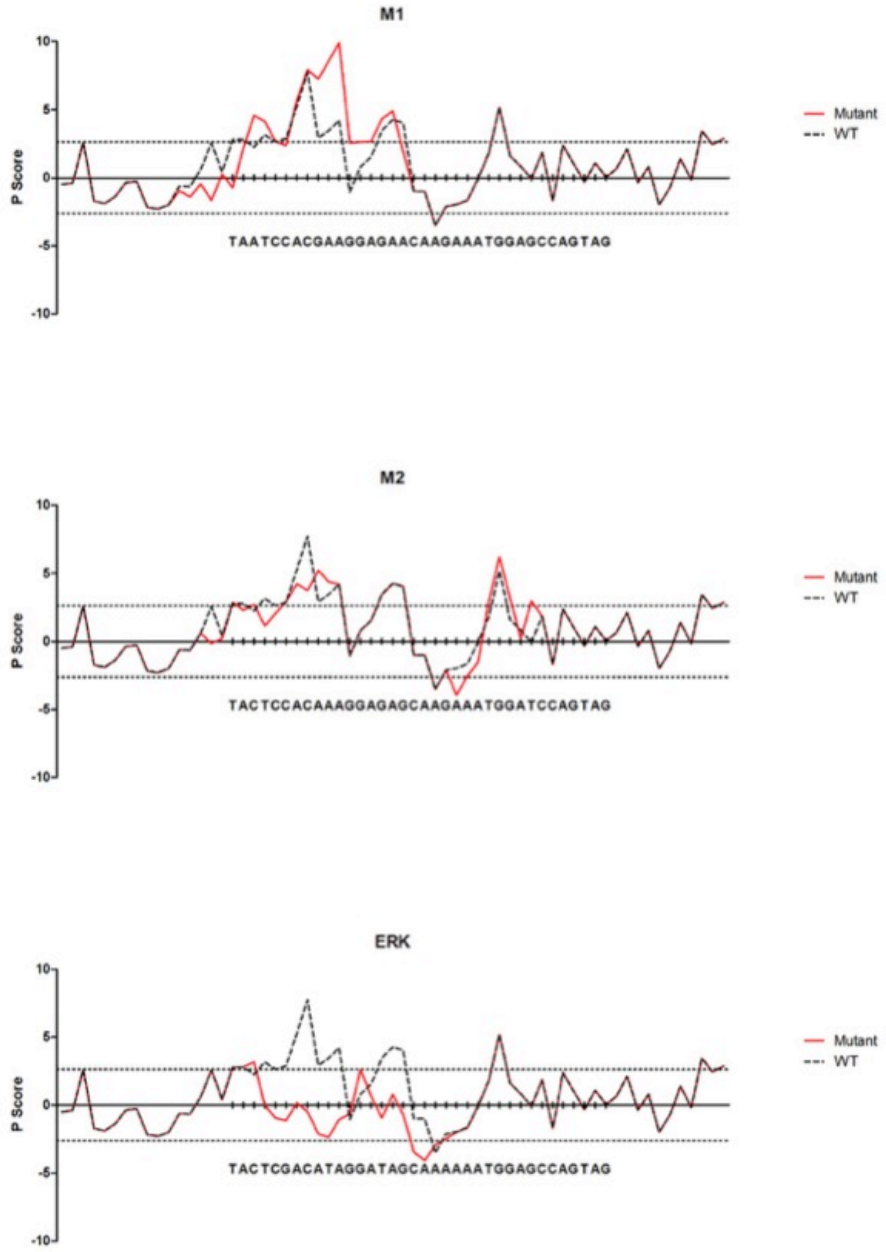


Figure 5.3 – Graphs showing predicted splice factor binding across the ESEtat region of the two selected mutants and the sequence identified by Erkelenz *et al.* compared to the wild type ESEtat P scores were calculated by PESX (Zhang and Chasin 2004) available from http:// cubio.biology.columbia.edu/pesx/pesx/. The software predicted increased splice factor binging in the M1 mutant, a small reduction in the M2 mutant and a severe defect in the ERK mutant

mutagenesis PCR. Sequences were confirmed by Sanger sequencing and then cloned into the original virus backbones.

To identify the *tat* mRNA isoforms produced by viruses with mutant ESEtat the plasmids were transfected into 293Tcells. RNA was extracted after 48 hours and converted to cDNA. PCR primers corresponding to position 640 and 5885 in HXB2 were used to amplify *tat* mRNA. Figure 5.4 shows the pattern of PCR products observed when examined on an agarose gel and the corresponding densitometry. The band sizes that would be expected for *tat*1, *tat*2, *tat*3 and *tat*4 are (210bp, 259bp, 283bp and 332bp respectively) the band observed at 598 bp was excised and sequenced and confirmed to be *vpr*. The densities of the bands were analysed by Image J to give an estimate of the quantity of DNA in each band. By normalising the amount of *tat* mRNA to *vpr* mRNA for each mutant it is possible to compare the relative abundance of *tat* transcripts. The M1 mutant shows a higher amount of tat1 transcripts compared to the wild type whereas both the M2 and ERK mutant showed reduced levels of *tat* mRNA production.

5.5 Detection of *tat* splice junctions by qPCR

Figure 5.2 shows the arrangement of exons required to make the four principal isoforms of *tat* mRNA. The junction of the major splice donor to the A3 splice site is unique to Tat1, the junction of the D2 splice donor to the A3 splice acceptor is unique to Tat2. As determination of mRNA levels by densitometry of gel electrophoresis is not the most accurate method quantitation was also carried out by qPCR. Primer and probe pairs were designed to amplify the unique splice junctions by designing primers which bind either side of the junction and a probe which binds to the unique sequence made by the joining of two exons. To confirm the specificity of the probes for spliced mRNA they were tested by including a reaction with plasmid DNA. No signal was observed from plasmid DNA or from no RT controls meaning that the probes specifically amplify the intended RNA sequences.

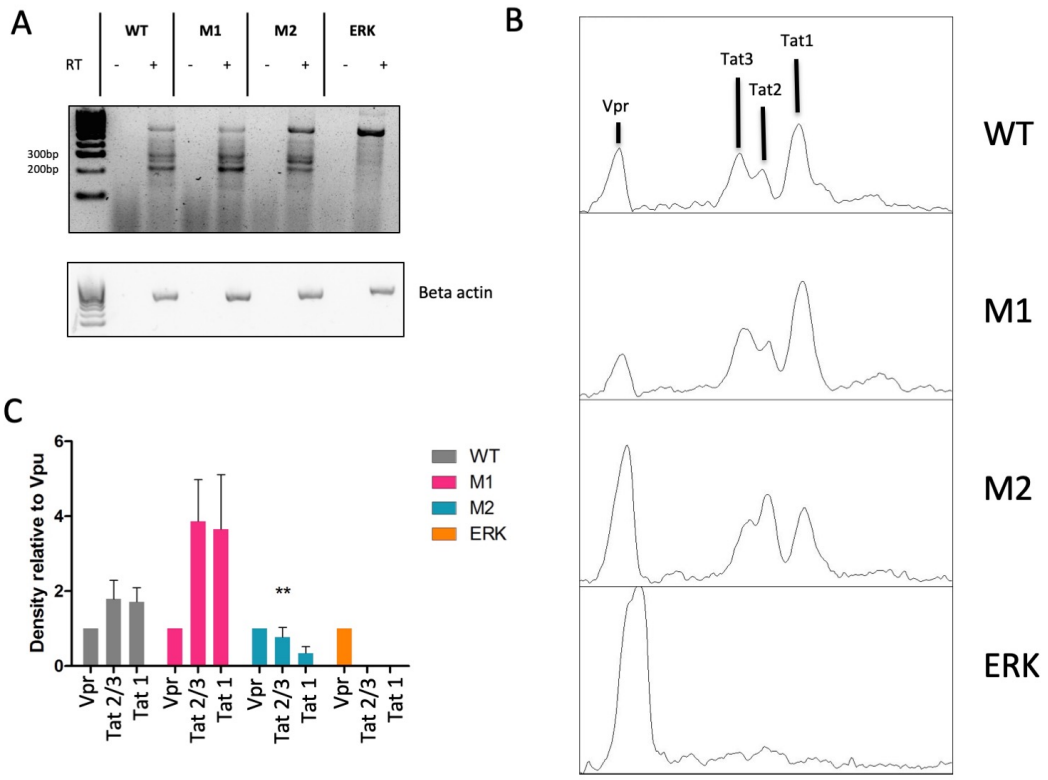


Figure 5.4 – **A.** Gel electrophoresis of Tat mRNA PCR products from 293T cells transfected with the ESEtat mutant plasmids. **B.** Example densitometry plots showing the observed bands and the corresponding HIV mRNA species. **C.** Results of three experiments showing the density of *tat* bands for each plasmid normalised to the *vpr* band. There was an apparent increase in *tat* mRNA in M1 compared to WT however this was not statistically significant. There was a reduction in observed *tat* mRNA in M2 compared to wild type (p=0.0048, two way ANOVA). Bands corresponding to *tat* mRNA were not detected from the ERK mutant

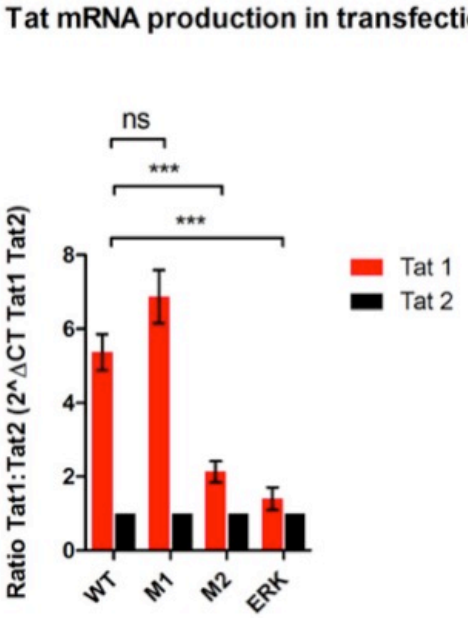


Figure 5.5 – qPCR analysis of *tat* mRNA isoforms. 293t cells were transfected with plasmids containing NL4.3GFP into which the ESEtat mutations were cloned. mRNA was harvested after 48 hours and reverse transcribed. Primers amplifying the D1A3 and D2A3 splice junctions were used to detect *tat1* and *tat2* mRNA species respectively. A difference in the ratio of tat1:tat2 production was seen in the M2 (p=0.004) and ERK mutants (p=0.002). Data represent the mean and standard error of three independent experiments.

The probes were used to evaluate whether there was altered use of *tat* splice sites in the mutants versus the wild type plasmid. The plasmids were transfected into 239T cells and RNA extracted 48 hours later. The *tat*1 signal was compared to the *tat*2 signal for each plasmid by calculating:

$$2^{\Delta CT}$$

where ΔCT is the CT value for *tat*1 minus the CT value *tat*2. Figure 5.5 shows a reduction in D1/A3 splice site use in the M2 and ERK mutants compared to the wild type. No significant difference was seen in the M1 mutant.

5.6 Effect of ESEtat mutants on silencing of HIV proviruses

A simple model system was used to explore the effect of the mutations on proviral silencing. The model was based on similar systems used to study HIV silencing in primary cells (Bosque and Planelles, 2011; Lassen *et al.*, 2012). The mutation containing plasmids based on NL4.3deltaENV.EGFP were used to generate a pseudotyped single round vector by co-transfecting the plasmids into 293T cells with a plasmid expressing vesicular stomatitis virus envelope glycoprotein (VSV-G). The vector was then used to infect Jurkat cells (Figure 5.6), after resting the cells for 72 hours the Jurkat cells can be thought of as having three different states; uninfected, infected with an expressing provirus or infected with a silenced provirus. Those harbouring an expressing provirus are GFP+ve without stimulation. Those that have silenced proviruses require additional stimulation to produce GFP. The cells are stimulated with PMA. The increase in the amount of GFP expressing cells gives an indication of how many proviruses were silent in the unstimulated cells. For example, if 10% of cells expressed GFP before stimulation and 20% post stimulation then at least half of all integration events were silenced and then induced by PMA. The cells were stimulated in the presence of 15nM efavirenz and 100nM raltegravir to prevent any residual viruses in the supernatant falsely increasing the GFP signal by newly infecting cells.

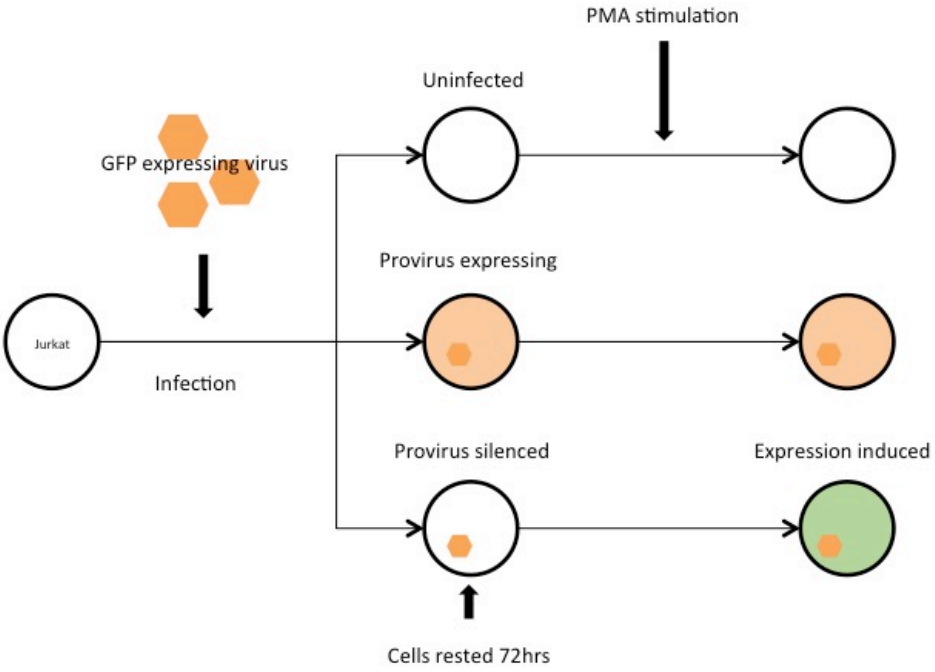


Figure 5.6 – Schematic of the silencing assay. After exposure to GFP expressing virus jurkat cells will be in one of three states; uninfected; infected and expressing GFP or infected with silencing of the provirus. After stimulation with PMA the silenced proviruses are made to produce GFP. The change in the proportion of GFP expressing jurkat cells gives a measure of the rate at which proviruses are silenced

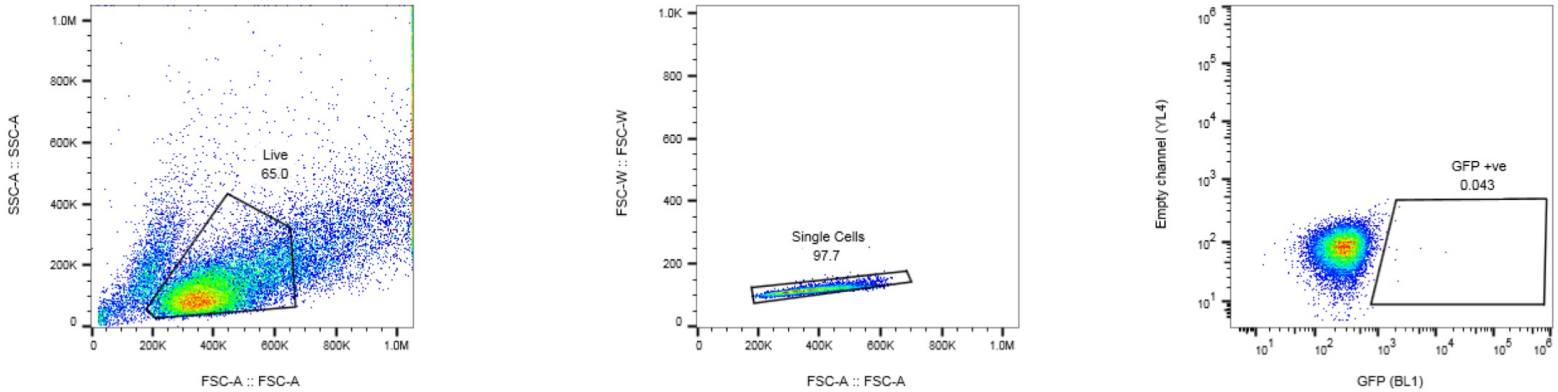


Figure 5.7 – Representative flow plot showing the gating strategy used to quantify the proportion of cells expressing GFP. After infection with GFP expressing virus live cells were identified by forward and side scatter characteristics (first panel), single events were isolated (second panel) and GFP expressing cells identified by comparison to an uninfected control group (third panel).

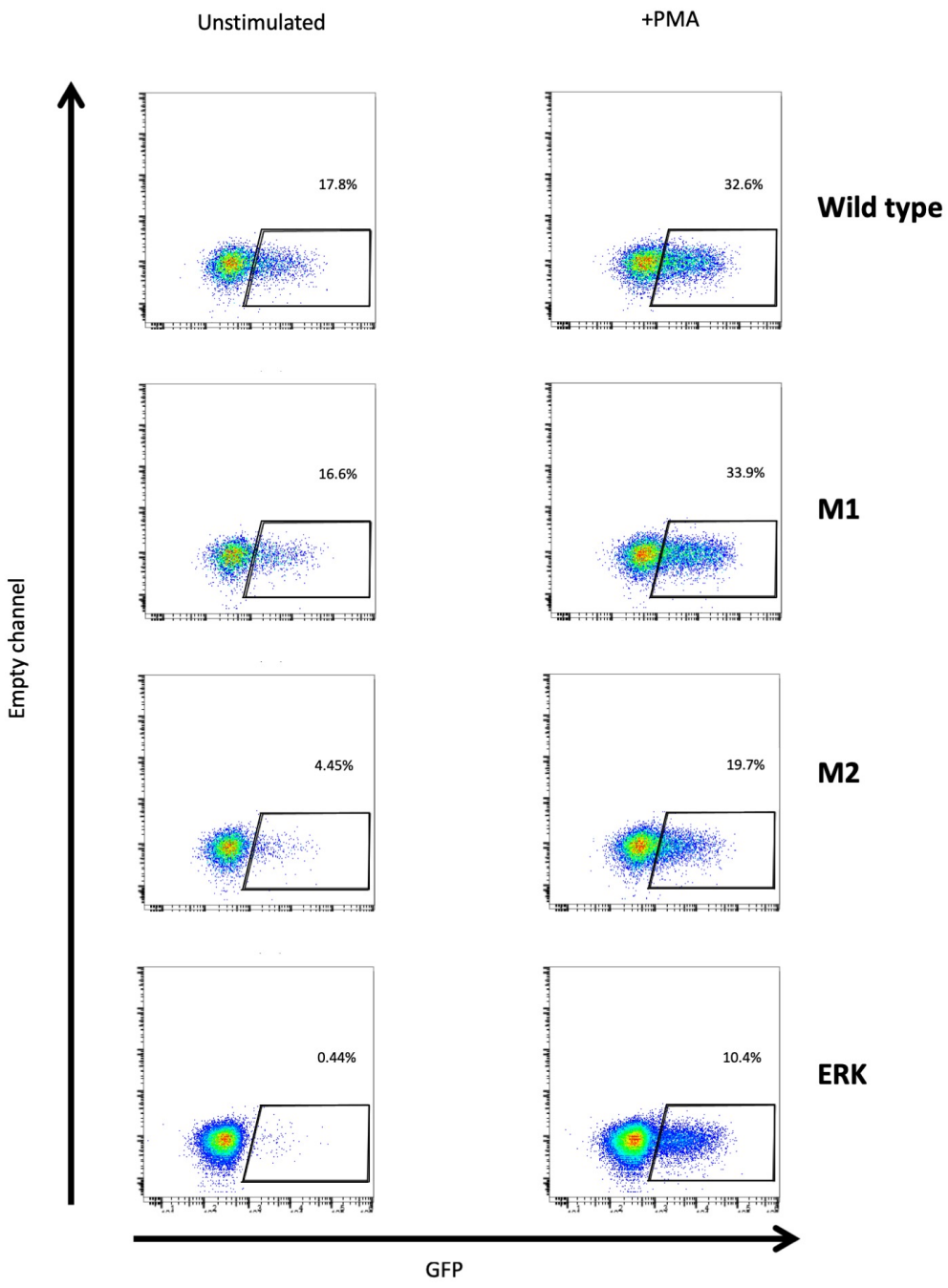


Figure 5.8 – Representative flow plots showing the increase in GFP expression following stimulation with PMA. $5x10^5$ Jurkat cells were infected with 50ng p24 NL4.3GFP; after 72 hours cells were stimulated with 200nM PMA in the presence of 100nM raltegravir and 15nM efavirenz. The cells were gated as shown in figure 5.6

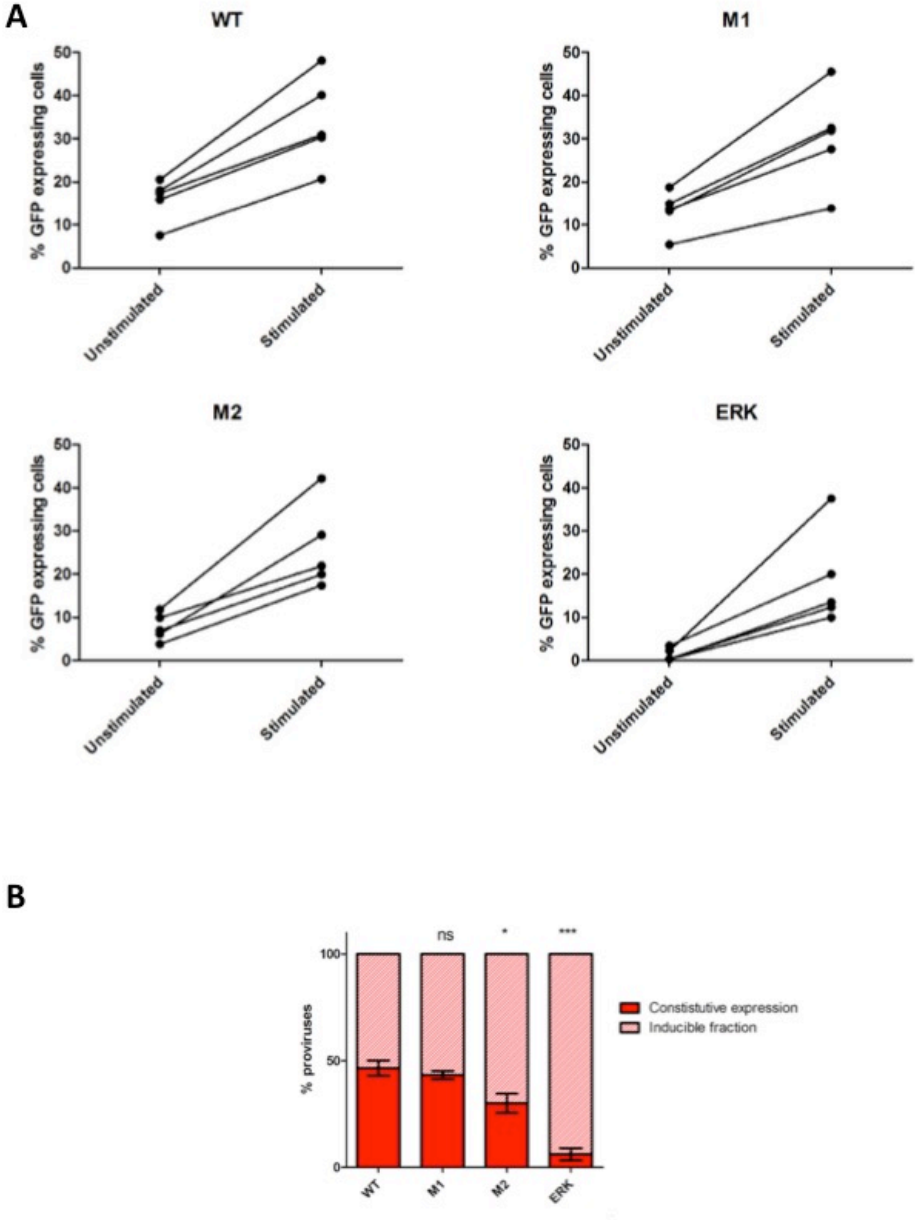


Figure 5.9 – Jurkat cells were infected with 50ng p24 per 500,000 cells of NL4.3GFP with wild type or mutant ESEtat regions. Cells were stimulated with 200nM PMA in the presence of 100nM raltegravir and 15nM efavirenz. Panel A shows the absolute % GFP expressing cells for the unstimulated baseline and stimulated conditions for each virus. Panel B shows the fraction of silenced proviruses derived from the fold change in GFP signal there was a statistically significant increase in silencing for the M2 mutant ($p=0.04$) and the ERK mutant ($p<0.001$) data points show mean and standard error, p values were determined by unpaired t test

The proportion of cells expressing GFP was quantified by flow cytometry (Figure 5.7). An uninfected control population was used to identify the forward and side scatter characteristics of live cells. Gating on forward scatter width versus area allowed identification of single events. Cells expressing GFP were identified by gating against the uninfected population. Figure 5.8 shows example flow plots obtained from unstimulated and PMA stimulated Jurkats infected with the viruses.

The M2 and ERK viruses displayed higher rates of silencing compared to the wild type (figure 5.9). The figure shows the changes in raw percentage GFP pre and post stimulation with PMA as well as the corresponding fold change in GFP signal for five independent experiments. The fold change in expression was 2.2 compared to 3.6 ($p=0.04$, two tailed t-test) for the M2 virus and 26 ($p<0.001$, two tailed t-test) for the ERK mutant. There was no significant difference between the wild type and M1 virus. The fold change figures correspond to a minimum estimate of the proportion of silenced proviruses of 52% for the wild type 72% for the M2 virus and 97% for the ERK mutant.

5.7 Splice junction use in the silencing model

mRNA was extracted from infected Jurkat cells as described. The splice junctions associated with *tat*1 and *tat*2 mRNA were detected by the primer-probe combinations described above. In addition to the *tat* specific primer-probe sets an additional probe that amplifies vpr was included to enable normalisation of transcript levels independent of the proportion of cells infected. Water only and no RT controls were included in the experiments as negative controls. Figure 5.10 shows the results; the bars were scaled so that the relative expression of *tat* mRNA seen in the wild type was set as 1. Compared to the wild type the M1 mutant showed an increased expression of *tat*1 and *tat*2 mRNA ($p=0.04$ and $p=0.005$, two tailed t-test) and the M2 mutant showed reduced expression ($p=0.01$ and $p=0.01$, two tailed t-test). *tat*1 and *tat*2 were not reliably detected from the ERK mutant despite good amplification of the *vpr* PCR.

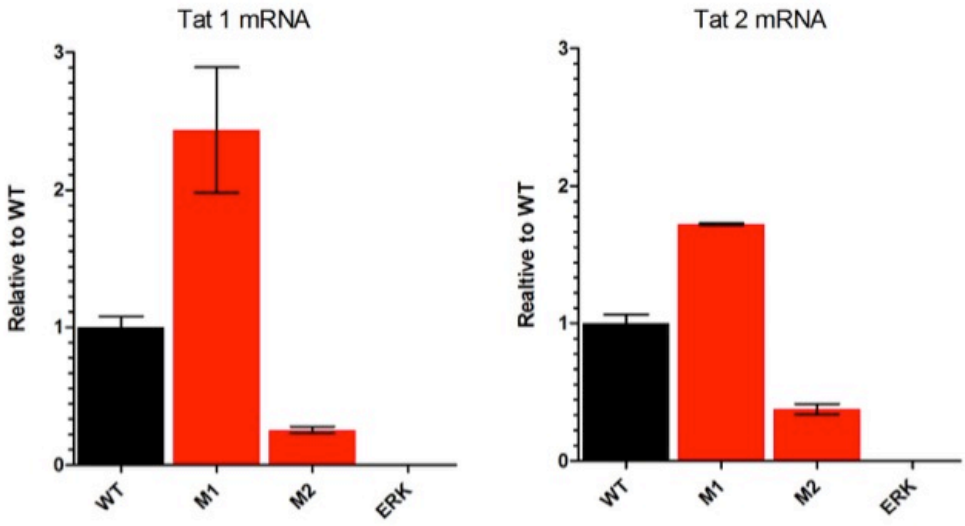


Figure 5.10 – Tat mRNA expression determined by qPCR. Jurkat cells were infected with GFP expressing virus. mRNA expression was determined by amplification of the D1A3 (tat1) or D2A3 (tat2) splice junctions. Abundance of mRNA was determined by comparison to a primer probe set amplifying *vpr* transcripts. Results were normalised to the wild type. M1 mutant showed an increased expression of Tat 1 and Tat 2 mRNA (p=0.04 and p=0.005) and the M2 mutant showed reduced expression (p=0.01 and p=0.01). Tat 1 and Tat 2 were not reliably detected from the ERK mutant despite good amplification of the *vpr* PCR. P values were determined using an unpaired t-test data points show mean and standard error.

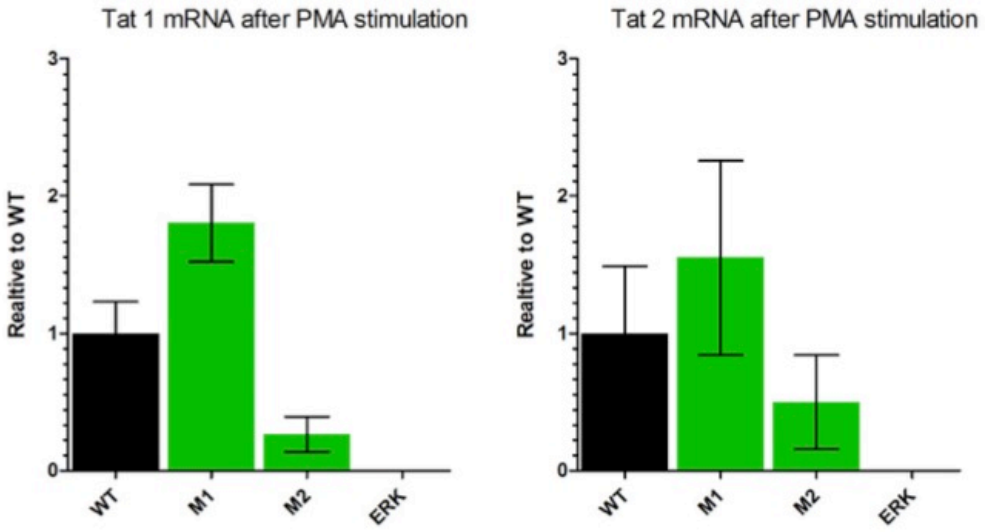


Figure 5.11 – Tat mRNA expression determined by qPCR as for figure 5.9. mRNA was extracted following stimulation with 200nM PMA, 100nM RAL and 15nM EFV. Data do not show a reversion to the wild type levels of tat mRNA production despite stimulation and despite the increase seen in GFP expression.

To evaluate whether the increase in GFP expression seen on stimulation was due to a reversion to the wild type splicing phenotype, mRNA was extracted from infected cells following stimulation with PMA in the presence of RAL and EFV (figure 5.11). A similar pattern of splicing was observed as for the unstimulated cells. There was no recovery in splice junction usage from the M2 or ERK mutants. The lack of recovery in the ERK mutants is particularly interesting as it suggests that the GFP signal produced in figure 5.9 is produced independently of production of Tat. Model systems have demonstrated that transcription for an LTR can proceed in the absence of Tat when an enhancer sequence from another gene is inserted upstream(West and Karn, 1999). In their paper West and Karn showed that the effect was independent of not only Tat but also of CDK9 suggesting that LTRs can be made to transcribe without activation of the positive transcriptional elongation complex.

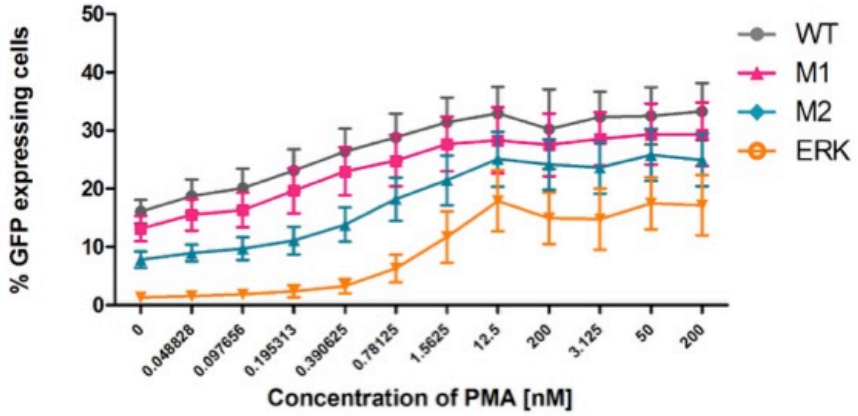
5.8 The effect of ESEtat mutants on reactivation threshold

The model system described above was used to evaluate whether mutations in ESEtat have an effect on the threshold required to reactive silenced proviruses. 50ng of p24 of pseudotyped virus was used to infect $5x10^5$ Jurkat cells. After 72 hours the cells were seeded into 12 wells of a 96 well plate. The media was removed from the cells and replaced with medium containing 100nM raltegravir and 15nM efavirenz as well as PMA in serial two fold dilutions starting at 200nM. An infected but unstimulated well was included in each experiment as a measure of baseline GFP expression. GFP expression was determined by flow cytometry using the method above. Figure 5.12a shows the mean %GFP expressing cells observed at each concentration of PMA for 5 experiments.

To detect a difference in reactivation threshold a dose response curve was fitted to the data using the software PRISM5. In order to fit the curve the data were scaled so that the baseline expression in the absence of PMA was set to 0 and the maximum observed GFP expression was set to 1, other values were scaled according to the formula:

$$GFP_{scaled} = \frac{GFP_{observed} - GFP_{min}}{GFP_{max} - GFP_{min}}$$

A



B

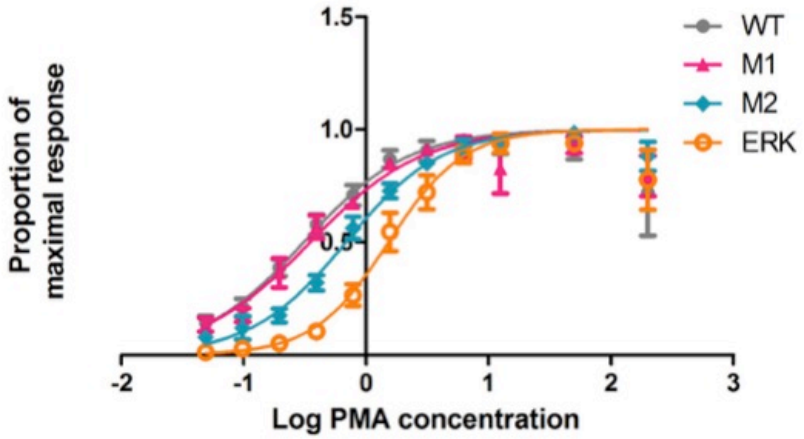


Figure 5.12 – Determination of reactivation threshold using PMA. Jurkat cells were infected with NL4.3GFP with mutant ESEtat. 72 hours after infection the cells were stimulated with PMA in serial two fold dilution in the presence of 100nM RAL and 15nM EFV. Panel A shows the % GFP expressing cells determined by flow cytometry 24 hours after stimulation with PMA. Panel B shows logarithmic curve fitting to the data. The maximal % GFP was set to 1 and the baseline GFP expression to 0. Other values of GFP expression were scaled accordingly. Data represent 5 independent experiments data points show mean and standard error.

	EC_{50}[nM]	95% Confidence interval	p-value
WT	0.3074	0.2263 to 0.4174	
M1	0.3472	0.2658 to 0.4534	p=0.63
M2	0.6797	0.5824 to 0.7931	p=0.025
ERK	1.517	1.226 to 1.877	p=0.0011

Table 5.3 – EC_{50} concentrations of PMA derived from the stimulations with serial dilutions of PMA. EC_{50}s and confidence intervals were determined using the logarithmic regression algorithm of Prism5. Confidence intervals were used to derive a T statistic and thus a p value for the comparison with the wild type. There was no significant difference in EC_{50} for the M1 mutant but the EC_{50}s were higher for both the M2 and ERK mutants

The analysis was used to compute the dose of PMA required to achieve 50% of maximum reactivation (EC_{50}).

Table 5.3 shows the EC_{50} for each mutant derived using the method above. There was no statistical difference between the WT and M1 viruses however the M2 virus had a significantly higher EC_{50} (0.6797nM vs 0.3074nM, p=0.025) as did the ERK mutant (1.517nM, p=0.0011) suggesting that these viruses have a higher threshold to be stimulated to reactivate. P values were derived by using the confidence interval obtained from the curve fitting to derive a T statistic for each comparison.

5.9 Response of ESEtat mutants to stimulation with latency reversing agents

The effect on reactivation by the latency reversing agents panobinostat and JQ1 was also studied. As described in more detail in chapters 1 and 3 panobinostat works by creating a histone environment which is more permissive by preventing the removal of acetyl groups from histones; JQ1 functions by increasing the pool of available PTEF-b to promote transcriptional elongation from the LTR.

The experimental design was identical to that described above. Infected cells were stimulated with serial two fold dilutions of panobinostat (PBST) starting from 400nM and JQ1 starting from 1000nM. To determine EC_{50}, curves were fitted to the resulting data as before. Figure 5.13 shows the curves fit to data obtained for stimulation with JQ1. Figure 5.14 shows the curves fit for the data obtained from stimulation with panobinostat. The WT and M1 viruses performed similarly for both sets of experiments. The M2 and ERK viruses had a higher threshold to reactivation with JQ1 compared to wild type (Table 5.4; 83.7nM vs 22.58nM and 256.8nM vs 22.58nM). There was a significant increase in the threshold for reactivation with panobinostat for the ERK mutant (Table 5.5; 12.86nM vs 1.6nM) there was a trend toward an increase in the M2 mutant but this was not statistically significant.

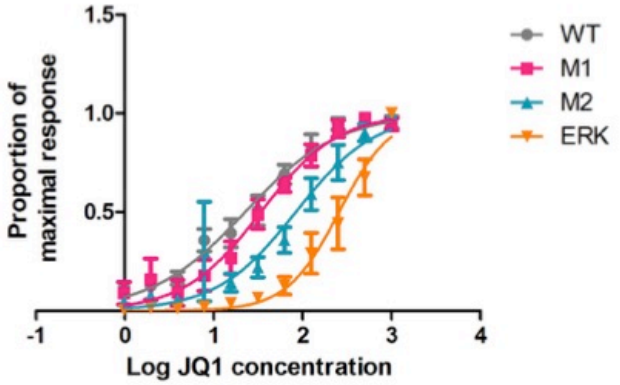


Figure 5.13 – Determination of activation threshold using JQ1. . Jurkat cells were infected with NL4.3GFP with mutant ESEtat. 72 hours after infection the cells were stimulated with JQ1 in serial two fold dilution in the presence of 100nM RAL and 15nM EFV. Curve fitting software (PRISM) was used to determine EC50 values

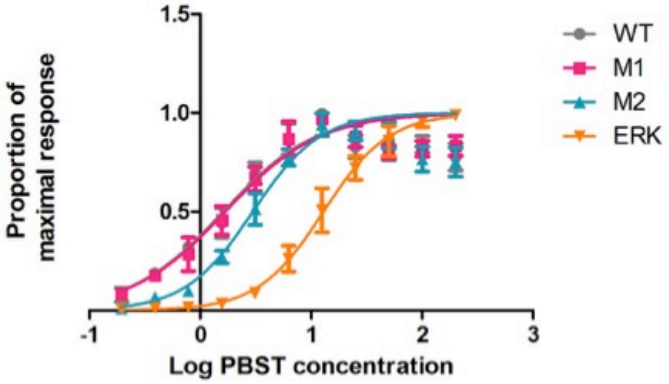


Figure 5.14 – Determination of activation threshold using panobinostat. . Jurkat cells were infected with NL4.3GFP with mutant ESEtat. 72 hours after infection the cells were stimulated with panobinostat in serial two fold dilution in the presence of 100nM RAL and 15nM EFV. Curve fitting software (PRISM) was used to determine EC50 values for each virus

	EC_{50}(JQ1 [nM])	95% Confidence interval	p-value
WT	22.58	18.09 to 28.17	
M1	33.88	26.43 to 43.43	p=0.11
M2	83.7	57.37 to 122.1	p=0.0258
ERK	256.8	211.3 to 312.1	p=0.0035

Table 5.4 – EC_{50} concentrations of JQ1 derived from the stimulations with serial dilutions. EC_{50}s and confidence intervals were determined using the curve fitting algorithm of Prism5. Confidence intervals were used to derive a T statistic and thus a p value for the comparison with the wild type. There was no significant difference in EC_{50} for the M1 mutant but the EC_{50}s were higher for both the M2 and ERK mutants

	EC_{50}(PBST [nM])	95% Confidence interval	p-value
WT	1.603	1.118 to 2.299	
M1	1.68	1.204 to 2.344	p=0.864
M2	2.972	2.226 to 3.969	p=0.11
ERK	12.86	11.01 to 15.03	p=0.0084

Table 5.5 – EC_{50} concentrations of PBST derived from the stimulations with serial dilutions. Confidence intervals were used to derive a T statistic and thus a p value for the comparison with the wild type. There was no significant difference in EC_{50} for the M1 mutant. There was a small but non significant difference in EC_{50} for the M2 mutant. The ERK mutant required a significantly higher concentration of panobinostat.

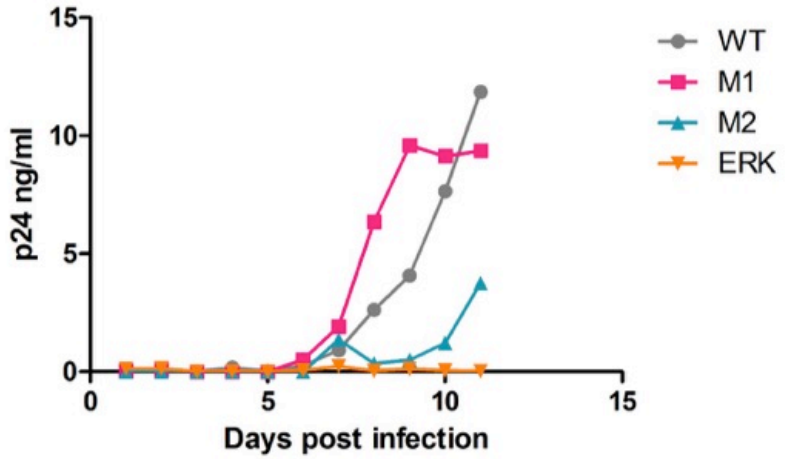
5.10 Experiments using replicating virus

As described previously the mutant ESEtat regions were also cloned into $HIV_{NL4.3}$ expressing a native envelope. $HIV_{NL4.3}$ is an X4 tropic virus that has been widely used as a standardised laboratory virus. As the virus expresses a native envelope it is replication competent. Virus was produced by transfecting 293T cells with plasmids containing the viral sequences. Viral supernatants were harvested and the p24 level measured to enable standardisation of p24 dose. SupT1-CCR5 cells were used in this set of experiments as they are the cells used in our laboratory to amplify reactivated latent virus from patient cells in the Viral Outgrowth Assay (Fun *et al.*, 2017). 5-8 million cells were infected with 15ng p24 per virus by spinoculation. Supernatant was sampled every 24 hours after infection and stored; at the end of the experiment supernatant p24 levels were determined by ELISA. Figure 5.15a shows an example of the outgrowth seen. The ERK mutant was not found to replicate in any experiment, confirming the results reported by Erkelenz *et al*.

Tiny variations in the number of cells seeded and quantity of p24 inoculated can have a significant effect on the absolute p24 level and on the day the virus is first detected. Therefore the rate of change in p24 was considered. P24 measurements were log transformed and fitted to a straight line (figure 5.15b). The slope of the line gives the log change in p24 per 24 hours. Lines were fit for data points with values 0.25ng/ml to 10ng/ml as this represents the dynamic range of the standard curves used to calibrate the ELISA results.

Figure 5.16 shows the mean log change per 24 hours for 5 experiments derived as described. There was no significant difference for the M1 mutant compared to wild type. The M2 mutant showed slightly reduced outgrowth (0.26 log/24hours vs 0.39log/24hours, p=0.04, two tailed t-test).

A



B

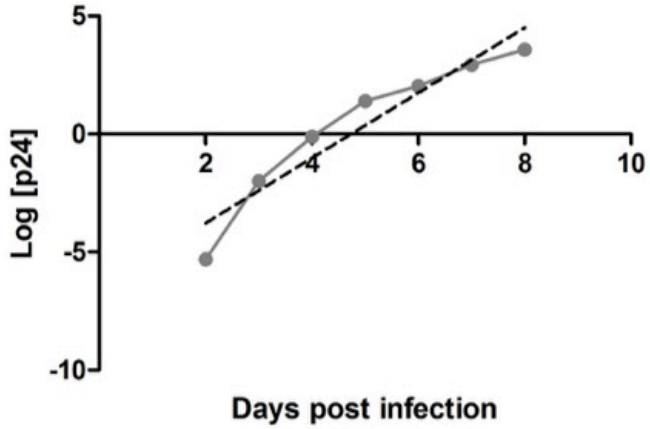


Figure 5.15 – Growth dynamics of ESEtat mutants in replicating virus. Panel A shows data from an example experiment. SupT1-CCR5 cells were infected with 15ng p24 of NL4.3 by spinoculation. After thorough washing cells were left in culture and supernatant harvested on a daily basis for p24 quantitation by ELISA. Panel B shows how the data were log transformed and analysed by linear regression. The slope of the line gives the log change in p24 per 24 hours. Note the ERK mutant did not grow out in any of these experiments

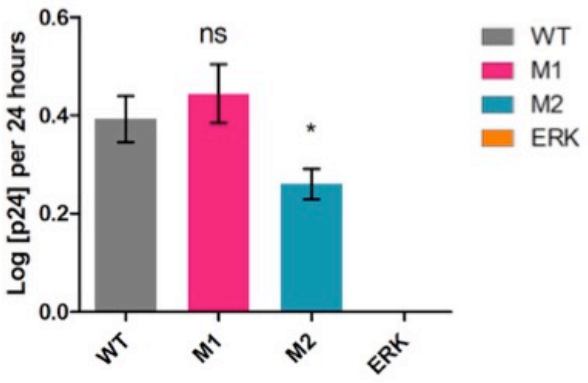


Figure 5.16 – The rate of growth of the replicating viruses with ESEtat mutations was evaluated by infecting SupT1CCR5 cells. The growth dynamics were analysed as shown in figure 5.14. M2 showed a significantly reduced rate of growth (p=0.04). Data are representative of 5 independent experiments and show mean with standard error bars p values were determined by unpaired t test

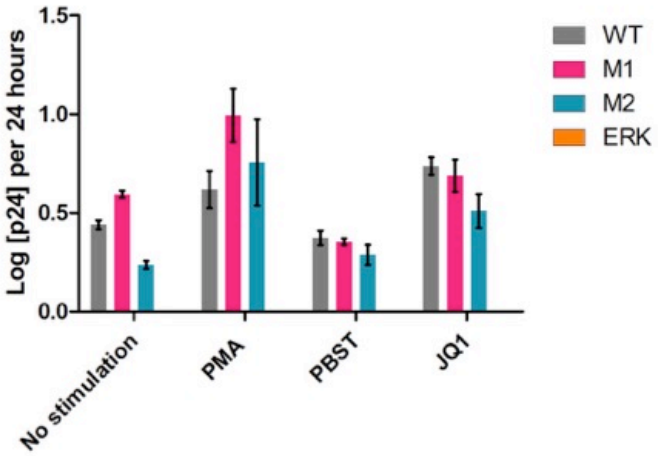


Figure 5.17 – The rate of growth of the replicating viruses with ESEtat mutations was evaluated by infecting $8x10^6$ SupT1CCR5 with 15ng of p24 cells in the presence of PMA 4nM, panobinostat 33nM and JQ1 650nM. The growth dynamics were analysed as shown in figure 5.14. Data show the mean and standard error of two independent experiments.

5.11 Additional stimulation in replicating virus

To examine whether growing out the virus in the presence of stimulating agents alters growth dynamics 8 million SupT1-CCR5 cells were infected with 15ng p24 of each of the four viruses. The culture was split after 24 hours into four subcultures which received either no stimulation, PMA, JQ1 or panobinostat. The concentration of each agent used was determined from the curves derived in sections 5.8 and 5.9. The dose of each agent required to achieve 80% response in the ERK mutant was used to ensure good activation of silenced proviruses without causing cell toxicity. Despite stimulation the ERK mutant could not be made to replicate. Stimulation with PMA and JQ1 appeared to increase the rate of p24 production for the WT, M1 and M2 viruses and recovered the M2 virus to at least the unstimulated WT levels. Intriguingly adding panobinostat to the culture medium did not lead to an increase in p24 growth rate. This suggests that chromatin mediated silencing of HIV is less important than the supply of transcription factors in determining the rate of virus outgrowth in this system.

5.12 Discussion and conclusions

These experiments have explored whether virus mediated phenomena can affect silencing behaviour. Critically the two mutations studied in these experiments were found in sequences isolated from circulating patient viruses meaning that viruses found *in vivo* can have mutations with the potential to alter the balance of actively transcribed and silenced proviruses.

Taken together with the wild type virus and the ERK mutant the four ESE_{tat} sequences studied can be thought of as belonging to a spectrum. The PESX scores for each sequence studied accurately predicted the difference in *tat* mRNA species observed (Figures 5.3, 5.4 and 5.10). The mutants which produced less *tat* mRNA (M2 and ERK) were observed to have a higher rate of proviral silencing. This is consistent with the idea that Tat production is essential to efficient transcription from the LTR and that a perturbation in this feedback loop enhances the

establishment of silencing. In some ways it is surprising that GFP production from the ERK virus could be induced by stimulation with PMA given that it produces negligible *tat* mRNA. This suggests that in this setting PMA is stimulating production of sufficient transcription factors to overcome the absence of Tat. This was confirmed by qPCR (figure 5.11) which showed *tat* mRNA production after stimulation did not increase to wild type levels.

ESE_{tat} mutations were also found to alter the threshold of silenced proviruses' response to stimulation. The mutations showing reduced splicing function also required the highest concentration of stimulating agents to induce transcription. It is interesting to note that despite PMA stimulation restoring GFP expression from the ERK mutant in the silencing model it was unable to restore replication in SupT1-CCR5 cells. This suggests that stimulation can induce protein production from an integrated ERK provirus but without ordered control of splicing the virus cannot replicate. Tat is not of itself required for the formation of infectious virions as evidenced by third generation lentiviral systems which function independently of Tat. Furthermore, Erkelenz *et al* showed that their virus could produce all other species of mRNA efficiently. These suggest that PMA, JQ1 or panobinostat mediated stimulation of the LTR are not sufficient to overcome the lack of Tat and restore replicative fitness.

The finding that minor mutations in ESE_{tat} can have profound changes in silencing behaviour has implications for our understanding of HIV latency *in vivo*. The data presented here suggest that viruses which have mutations affecting the Tat feedback loop need a higher level of stimulation to achieve the same level of transcription from the LTR. This is one possible explanation for findings such as those of Ho *et al*, who reported that a proportion of latent proviruses are not reactivated despite having a full length genome with no significant deletions (Ho *et al.*, 2013). Critically the M2 mutant in this study was able to replicate in culture, similarly the 'non induced' proviruses of the Ho study could replicate when their sequences were reconstructed.

The observation that the M2 mutant was replication competent but with a reduced rate of p24 production is also significant. In the standard virus outgrowth assay for determining the size of the replication competent latent HIV reservoir, determination of whether or not a virus has been released from a latently infected cell is made by detecting p24 in the supernatant after the virus has had time to replicate in amplifying cells. A virus which replicated less efficiently in the culture could give rise to a false negative if it has not produced enough p24 to cross the detection threshold by the day that the supernatant is harvested.

The combined error rate of HIV reverse transcriptase and cellular Pol II has been observed to be $4x10^{-3}$ per nucleotide per replication cycle (Roberts, Bebenek and Kunkel, 1988; Cuevas *et al.*, 2015) meaning that at least one single nucleotide mutation is expected after each replication cycle. This explains the observation that the latent reservoir *in vivo* consists of similar but non-identical viruses archived from the founder sequence onwards(Bailey *et al.*, 2006; Oliveira *et al.*, 2017). It is therefore not inconceivable that among the total archived viruses within a host some have mutations altering the response to latency reversing strategies. This implies that latency reversal employed as part of a kick and kill strategy for HIV cure could result in preferential activation of a proportion of proviruses, leaving a residual reservoir of progressively more reactivation resistant archived viruses. As efforts to cure HIV progress it will be informative to explore whether a background of hard to reactivate viruses influences the dynamics of reservoir decay. Furthermore it is possible that within a patient on suppressive therapy the most easily reactivated viruses are spontaneously activated and eliminated while harder to reactivate viruses remain. This hypothesis could be tested by examining the time taken for viruses to be identified in viral outgrowth assays from those who have had a suppressed viral load for a long time and compare these to more recently treated patients.

As noted in the introduction a virally encoded enhancer of latency appears to be counterproductive for the virus however it would be consistent with the modelling work of Leor Weinberger (Rouzine, Weinberger and Weinberger, 2015) that suggests promoting latency during early mucosal infection enables HIV to

evade detection by the immune system and survive to establish itself in the blood compartment. As our understanding of how latency is beneficial to retroviruses in the absence of antiretroviral therapy deepens it is possible that more examples of viruses manipulating the transcriptional environment to promote latency may emerge.

It is important to acknowledge some of the shortcomings in the experimental approach used in this chapter. One is that the effect of the mutations was examined in bulk. One way in which this could have been addressed would have been to sort cells to create a system analogous to the j-lat model using viruses with mutant ESEtat regions and growing out clones of singly infected cells. The problem with this approach would have been that the effect of a single integration site could predominate rather than the viral effect under study. The bulk approach has the advantage of minimising the contribution of integration sites to the signal observed although it cannot exclude an altered distribution of integration sites in the mutant viruses. The growth of a cell clone would also have the advantage of excluding preintegration DNA from the signal. Despite using RAL and EFV in the culture medium in the stimulation studies it is possible that GFP could be expressed from unintegrated DNA although it could be expected that this problem would affect all the viruses equally and would still be dependent on efficient production of Tat.

An effect on Tat protein production was not demonstrated here. It is possible that although the mutations clearly have a consistent effect on *tat* mRNA production and splicing this effect is not responsible for the observed differences between the viruses. It is worth noting however that Erkelenz *et al.* did in fact demonstrate such a reduction.

The outgrowth test used here was not the ideal way to demonstrate an effect on viral fitness. Ideally viruses should be seeded into the same flask in a range of ratios (e.g. A:B = 90:10, 80:20, 70:30, 60:40 50:50 etc.) after a defined period the supernatant is sampled and any virus sequenced. The viral sequence recovered identifies the virus that has out-competed the other. One problem with this

approach is that as the difference between viruses becomes smaller the more mixes closer to 50:50 are required to identify the difference. Another problem is that the number of experiments becomes unwieldy, particularly when a four way comparison is required.

Chapter 6: Summary and concluding thoughts

HIV infection remains a global public health problem. The present treatment involving patients taking life long suppressive medication has achieved impressive results in terms of quality of life and prognosis for those who are able to access antiretroviral therapy. Nonetheless the global prevalence of this infection has continued to increase, driven by a combination of new infections and increased survival of those infected. The economic burden in countries with large populations of people living with HIV is considerable. The ability to eradicate the virus from an individual would be a powerful tool in reversing this situation.

As discussed in Chapter 1 there is reason to think that the principal barrier to a cure are latently infected CD4+ T cells. In this thesis the mechanisms maintaining latency as well as factors which influence the efficiency of waking virus have been examined. In Chapter 3 a cell line model of latency was used to examine how clones of latently infected cells harbouring a provirus in an identical integration site might respond to agents designed to provoke transcription and reverse latency. It was observed that only a proportion of cells within a clone respond to each stimulation and that this maximum proportion is fixed for each integration site and activating agent combination. In addition it was possible to show that these fixed upper limits could be overcome by utilising combinations of latency reversing agents. These observations provide insights into how best to shrink the reservoir of latently infected cells; particularly in light of emerging evidence of clonal proliferation of infected cells *in vivo.*

In Chapter 4 the potential role of a recently described protein complex (HUSH) in promoting latency was examined. This yielded mixed results; it was not possible to demonstrate that HUSH had a significant contribution to the silencing of most clones in the J-lat model of latency. Some of the data here suggest that HUSH may have a role in early establishment of silencing however the results were not able to demonstrate this conclusively. Taken together with the results reported by the

Lehner lab HUSH may be involved in specific contexts which were not fully explored in this chapter.

In Chapter 5 the contribution of viral factors to the silencing of HIV proviruses was examined. Polymorphisms in a recently described splice enhancer were used as a model to examine the paradigm that variations in the Tat-TAR axis driving HIV transcription could influence the rate of silencing. Mutations which had a deleterious effect on splicing Tat mRNA were observed to promote HIV silencing while only having a moderate effect on the ability of the virus to replicate. These mutations also altered the threshold at which stimulation would induce transcription from a silenced provirus. This has important consequences for understanding how the reservoir is structured in vivo and suggests that some cells may harbour viruses requiring greater stimulation to induce expression. This in turn may provide some explanation for the observation reported by the Siliciano group that some cells *in vivo* harbour apparently intact proviruses from which transcription cannot be induced.

The data here also suggest avenues for future investigation especially in examining cells *ex vivo*. Data concerning the behaviour of clonally integrated proviruses in patients is scarce. Were it possible to identify and maintain a clone from a patient than it would be instructive to replicate the experiments here. A an alternate approach would be to take a longitudinal look at changes in the phylogeny of the HIV reservoir in patients receiving a HDACi or other latency reversing agent. A differential change in the representation of one clone over another would suggest that there is a corresponding difference in sensitivity to latency reversal between to two clones.

Further work on viral factors contributing to latency is needed. An obvious starting point would be to examine how polymorphisms in the LTR and in the coding of Tat affect transcription and silencing in a similar way to the ESEtat mutants studied here. Another approach would be a comparison of the sequences of viruses successfully reactivated from primary cells to those of intact non-induced proviruses to identify sequences which are over represented in the latent

viruses. Clearly this approach would require a large data set however the rapid pace of advances in sequencing technology in combination with the ongoing interest in publishing sets of these sequences will make this comparison feasible in the near future.

References

Adelman, K. and Lis, J. T. (2012) 'Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans', *Nature Reviews Genetics*. Nature Publishing Group, 13(10), pp. 720–731. doi: 10.1038/nrg3293.

Anderson, J. A. *et al.* (2011) 'Clonal sequences recovered from plasma from patients with residual HIV-1 viremia and on intensified antiretroviral therapy are identical to replicating viral RNAs recovered from circulating resting CD4+ T cells', *Journal of Virology*, 85(10), pp. 5220–5223.

Archin, N. *et al.* (2008) 'Valproic acid without intensified antiviral therapy has limited impact on persistent HIV infection of resting CD4+T cells', *AIDS*, 22(10), pp. 1131–1135.

Archin, N. M., Espeseth, A., *et al.* (2009) 'Expression of latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid.', *AIDS research and human retroviruses*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, 25(2), pp. 207–12. doi: 10.1089/aid.2008.0191.

Archin, N. M., Keedy, K. S., *et al.* (2009) 'Expression of latent human immunodeficiency type 1 is induced by novel and selective histone deacetylase inhibitors.', *AIDS (London, England)*, 23(14), pp. 1799–806. doi: 10.1097/QAD.0b013e32832ec1dc.

Archin, N. M. *et al.* (2012) 'Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy.', *Nature*, 487(7408), pp. 482–5. doi: 10.1038/nature11286.

Archin, N. M. *et al.* (2014) 'HIV-1 expression within resting CD4+ T cells after multiple doses of vorinostat.', *The Journal of infectious diseases*, 210(5), pp. 728–35. doi: 10.1093/infdis/jiu155.

Bailey, J. R. *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(13), pp. 6441–57. doi: 10.1128/JVI.00591-06.

Banerjee, C. *et al.* (2012) 'BET bromodomain inhibition as a novel strategy for reactivation of HIV-1.', *Journal of leukocyte biology*, 92(6), pp. 1147–54. doi: 10.1189/jlb.0312165.

Bannister, A. J. and Kouzarides, T. (2005) 'Reversing histone methylation', *Nature*. Nature Publishing Group, 436(7054), pp. 1103–1106. doi: 10.1038/nature04048.

Barboric, M. *et al.* (2007) 'Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription.', *Nucleic acids research*, 35(6), pp. 2003–12. doi: 10.1093/nar/gkm063.

Barnett, S. W. *et al.* (1991) 'Characterization of human immunodeficiency virus type 1 strains recovered from the bowel of infected individuals.', *Virology*, 182(2), pp. 802–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2024498 (Accessed: 12 January 2016).

Barré-Sinoussi, F. *et al.* (1983) 'Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).', *Science (New York, N.Y.)*, 220(4599), pp. 868–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6189183 (Accessed: 14 November 2014).

Benkirane, M. *et al.* (1998) 'Activation of integrated provirus requires histone acetyltransferase: p300 and P/CAF are coactivators for HIV-1 Tat', *Journal of Biological Chemistry*, 273(38), pp. 24898–24905. doi: 10.1074/jbc.273.38.24898.

Besnard, E. *et al.* (2016) 'The mTOR Complex Controls HIV Latency', *Cell Host & Microbe*, 20(6), pp. 785–797. doi: 10.1016/j.chom.2016.11.001.
Besson, G. J. *et al.* (2014) 'HIV-1 DNA Decay Dynamics in Blood During More Than a Decade of Suppressive Antiretroviral Therapy', *Clinical Infectious Diseases*. Oxford University Press, 59(9), pp. 1312–1321. doi: 10.1093/cid/ciu585.
Binka, M. *et al.* (2012) 'The Activity Spectrum of Vif from Multiple HIV-1 Subtypes against APOBEC3G, APOBEC3F, and APOBEC3H', *Journal of Virology*, 86(1), pp. 49–59. doi: 10.1128/JVI.06082-11.
Bisgrove, D. A. *et al.* (2007) 'Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription.', *Proceedings of the National Academy of Sciences of the United States of America*, 104(34), pp. 13690–5. doi: 10.1073/pnas.0705053104.
Blackhall, F. H. *et al.* (2001) 'A phase II trial of bryostatin 1 in patients with non-Hodgkin's lymphoma', *British Journal of Cancer*. Nature Publishing Group, 84(4), pp. 465–469. doi: 10.1054/bjoc.2000.1624.
Blankson, J. N. *et al.* (2007) 'Isolation and Characterization of Replication-Competent Human Immunodeficiency Virus Type 1 from a Subset of Elite Suppressors', *Journal of Virology*, 81(5), pp. 2508–2518. doi: 10.1128/JVI.02165-06.
Blazkova, J. *et al.* (2012) 'Paucity of HIV DNA methylation in latently infected, resting CD4+ T cells from infected individuals receiving antiretroviral therapy.', *Journal of virology*. American Society for Microbiology (ASM), 86(9), pp. 5390–2. doi: 10.1128/JVI.00040-12.
Boehm, D. *et al.* (2013) 'BET bromodomain-targeting compounds reactivate HIV from latency via a Tat-independent mechanism.', *Cell cycle (Georgetown, Tex.)*, 12(3), pp. 452–62. doi: 10.4161/cc.23309.
Bosque, A. and Planelles, V. (2009) 'Induction of HIV-1 latency and reactivation in primary memory CD4+ T cells', *Blood*, 113(1), pp. 58–65. doi: 10.1182/blood-2008-07-168393.
Bosque, A. and Planelles, V. (2011) 'Studies of HIV-1 latency in an ex vivo model that uses primary central memory T cells.', *Methods (San Diego, Calif.)*, 53(1), pp. 54–61. doi: 10.1016/j.ymeth.2010.10.002.
Bouchat, S. *et al.* (2012) 'Histone methyltransferase inhibitors induce HIV-1 recovery in resting CD4+ T cells from HIV-1-infected HAART-treated patients', *AIDS*, 26(12), pp. 1473–1482. doi: 10.1097/QAD.0b013e32835535f5.
Brady, T. *et al.* (2009) 'HIV integration site distributions in resting and activated CD4+ T cells infected in culture.', *AIDS (London, England)*, 23(12), pp. 1461–71. doi: 10.1097/QAD.0b013e32832caf28.
Brennan, T. P. *et al.* (2009) 'Analysis of human immunodeficiency virus type 1 viremia and provirus in resting CD4+ T cells reveals a novel source of residual viremia in patients on antiretroviral therapy.', *Journal of virology*, 83(17), pp. 8470–81. doi: 10.1128/JVI.02568-08.
Brown, C. E. *et al.* (2000) 'The many HATs of transcription coactivators', *Trends in Biochemical Sciences*, 25(1), pp. 15–19. doi: 10.1016/S0968-0004(99)01516-9.
Bui, J. K. *et al.* (2017) 'Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir', *PLOS Pathogens*. Edited by S. R. Ross. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 13(3), p. e1006283. doi: 10.1371/journal.ppat.1006283.
Buratowski, S. (2003) 'The CTD code', *Nature Structural & Molecular Biology*. Nature Publishing Group, 10(9), pp. 679–680. doi: 10.1038/nsb0903-679.

Buzon, M. J. *et al.* (2014) 'HIV-1 persistence in CD4+ T cells with stem cell–like properties', *Nature Medicine*, 20(2), pp. 139–142. doi: 10.1038/nm.3445.
Byvoet, P. *et al.* (1972) 'The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells.', *Archives of biochemistry and biophysics*, 148(2), pp. 558–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/5063076 (Accessed: 24 February 2017).
Campbell-Yesufu, O. T. and Gandhi, R. T. (2011) 'Update on human immunodeficiency virus (HIV)-2 infection.', *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 52(6), pp. 780–7. doi: 10.1093/cid/ciq248.
Cao, Y. *et al.* (1995) 'Virologic and Immunologic Characterization of Long-Term Survivors of Human Immunodeficiency Virus Type 1 Infection', *New England Journal of Medicine*, 332(4), pp. 201–208. doi: 10.1056/NEJM199501263320401.
Chan, C. N. *et al.* (2016) 'HIV-1 latency and virus production from unintegrated genomes following direct infection of resting CD4 T cells.', *Retrovirology*. BioMed Central Ltd, 13(1), p. 1. doi: 10.1186/s12977-015-0234-9.
Charlins, P. *et al.* (2017) 'A humanized mouse-based HIV-1 viral outgrowth assay with higher sensitivity than in vitro qVOA in detecting latently infected cells from individuals on ART with undetectable viral loads', *Virology*, 507, pp. 135–139. doi: 10.1016/j.virol.2017.04.011.
Chavez, L., Calvanese, V. and Verdin, E. (2015) 'HIV Latency Is Established Directly and Early in Both Resting and Activated Primary CD4 T Cells.', *PLoS pathogens*. Public Library of Science, 11(6), p. e1004955. doi: 10.1371/journal.ppat.1004955.
du Chéné, I. *et al.* (2007) 'Suv39H1 and HP1gamma are responsible for chromatin-mediated HIV-1 transcriptional silencing and post-integration latency.', *The EMBO journal*. European Molecular Biology Organization, 26(2), pp. 424–35. doi: 10.1038/sj.emboj.7601517.
Cherblanc, F. L. *et al.* (2013) 'Chaetocin is a nonspecific inhibitor of histone lysine methyltransferases', *Nature Chemical Biology*. Nature Research, 9(3), pp. 136–137. doi: 10.1038/nchembio.1187.
Chomont, N. *et al.* (2009) 'HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation.', *Nature medicine*. Nature Publishing Group, 15(8), pp. 893–900. doi: 10.1038/nm.1972.
Choudhary, S. K., Archin, N. M. and Margolis, D. M. (2008) 'Hexamethylbisacetamide and disruption of human immunodeficiency virus type 1 latency in CD4(+) T cells.', *The Journal of infectious diseases*. Oxford University Press, 197(8), pp. 1162–70. doi: 10.1086/529525.
Christ, F. and Debyser, Z. (2013) 'The LEDGF/p75 integrase interaction, a novel target for anti-HIV therapy.', *Virology*, 435(1), pp. 102–9. doi: 10.1016/j.virol.2012.09.033.
Chun, T.-W. *et al.* (1997) 'Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection', *Nature*, 387(6629), pp. 183–188.
Chun, T. *et al.* (1997) 'Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection', *Nature*, 387(5429), pp. 183–187. doi: 10.1038/246170a0.
Chun, T. W. *et al.* (1997) 'Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy.', *Proceedings of the National Academy of Sciences of the United States of America*, 94(24), pp. 13193–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=24285&tool=pmcentrez&rendertype=abstract (Accessed: 8 December 2015).

Clapham, P. R. (2001) 'HIV-1 receptors and cell tropism', *British Medical Bulletin*, 58(1), pp. 43–59. doi: 10.1093/bmb/58.1.43.
Cohen, M. S. *et al.* (2011) 'Prevention of HIV-1 infection with early antiretroviral therapy.', *The New England journal of medicine*, 365(6), pp. 493–505. doi: 10.1056/NEJMoa1105243.
Cohn, L. B. *et al.* (2015) 'HIV-1 integration landscape during latent and active infection.', *Cell*. Elsevier, 160(3), pp. 420–32. doi: 10.1016/j.cell.2015.01.020.
Contreras, X. *et al.* (2007) 'HMBA releases P-TEFb from HEXIM1 and 7SK snRNA via PI3K/Akt and activates HIV transcription.', *PLoS pathogens*. Edited by J. A. T. Young. Public Library of Science, 3(10), pp. 1459–69. doi: 10.1371/journal.ppat.0030146.
Core, L. J. and Lis, J. T. (2008) 'Transcription regulation through promoter-proximal pausing of RNA polymerase II.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 319(5871), pp. 1791–2. doi: 10.1126/science.1150843.
Craigie, R. and Bushman, F. D. (2012) 'HIV DNA integration.', *Cold Spring Harbor perspectives in medicine*, 2(7), p. a006890. doi: 10.1101/cshperspect.a006890.
Crooks, A. M. *et al.* (2015) 'Precise Quantitation of the Latent HIV-1 Reservoir: Implications for Eradication Strategies.', *The Journal of infectious diseases*, p. jiv218-. doi: 10.1093/infdis/jiv218.
Cuevas, J. M. *et al.* (2015) 'Extremely High Mutation Rate of HIV-1 In Vivo', *PLOS Biology*. Edited by S. L. Rowland-Jones. Public Library of Science, 13(9), p. e1002251. doi: 10.1371/journal.pbio.1002251.
Dahabieh, M. S. *et al.* (2013) 'A double-fluorescent HIV-1 reporter shows that the majority of integrated HIV-1 is latent shortly after infection.', *Journal of virology*, 87(8), pp. 4716–4727. doi: 10.1128/JVI.03478-12.
Dar, R. D. *et al.* (2014) 'Screening for noise in gene expression identifies drug synergies.', *Science (New York, N.Y.)*, 344(6190), pp. 1392–6. doi: 10.1126/science.1250220.
Darcis, G. *et al.* (2015) 'An In-Depth Comparison of Latency-Reversing Agent Combinations in Various In Vitro and Ex Vivo HIV-1 Latency Models Identified Bryostatin-1+JQ1 and Ingenol-B+JQ1 to Potently Reactivate Viral Gene Expression', *PLOS Pathogens*. Edited by J. Karn. Public Library of Science, 11(7), p. e1005063. doi: 10.1371/journal.ppat.1005063.
Debiaggi, M. *et al.* (2000) 'Quantification of HIV-1 proviral DNA in patients with undetectable plasma viremia over long-term highly active antiretroviral therapy.', *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*, 4(4), pp. 187–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11231180 (Accessed: 13 January 2015).
Deeks, S. G. (2012) 'HIV: Shock and kill.', *Nature*. Nature Publishing Group, 487(7408), pp. 439–40. doi: 10.1038/487439a.
Deng, K. *et al.* (2015) 'Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations.', *Nature*. Nature Publishing Group. doi: 10.1038/nature14053.
Deng, L. *et al.* (2000) 'Acetylation of HIV-1 Tat by CBP/P300 Increases Transcription of Integrated HIV-1 Genome and Enhances Binding to Core Histones', *Virology*, 277(2), pp. 278–295. doi: 10.1006/viro.2000.0593.
Díaz, L. *et al.* (2015) 'Bryostatin activates HIV-1 latent expression in human astrocytes through a PKC and NF-κB-dependent mechanism.', *Scientific reports*.

Nature Publishing Group, 5, p. 12442. doi: 10.1038/srep12442.
Dokmanovic, M., Clarke, C. and Marks, P. A. (2007) 'Histone deacetylase inhibitors: overview and perspectives.', *Molecular cancer research : MCR*. American Association for Cancer Research, 5(10), pp. 981–9. doi: 10.1158/1541-7786.MCR-07-0324.
Dolin, R. *et al.* (1991) 'The Safety and Immunogenicity of a Human Immunodeficiency Virus Type 1 (HIV-1) Recombinant gp160 Candidate Vaccine in Humans', *Annals of Internal Medicine*. American College of Physicians, 114(2), pp. 119–127. doi: 10.7326/0003-4819-114-2-119.
Dosenovic, P. *et al.* (2015) 'Immunization for HIV-1 Broadly Neutralizing Antibodies in Human Ig Knockin Mice.', *Cell*, 161(7), pp. 1505–15. doi: 10.1016/j.cell.2015.06.003.
Van Duyne, R. *et al.* (2008) 'Lysine methylation of HIV-1 Tat regulates transcriptional activity of the viral LTR', *Retrovirology*. BioMed Central, 5(1), p. 40. doi: 10.1186/1742-4690-5-40.
Elliott, J. H. *et al.* (2015) 'Short-term administration of disulfiram for reversal of latent HIV infection: a phase 2 dose-escalation study', *The Lancet HIV*. Elsevier, 2(12), pp. e520–e529. doi: 10.1016/S2352-3018(15)00226-X.
Emery, A. *et al.* (2017) 'Characterizing HIV-1 Splicing by Using Next-Generation Sequencing', *Journal of Virology*. American Society for Microbiology, 91(6), pp. e02515-16. doi: 10.1128/JVI.02515-16.
Emiliani, S. *et al.* (1996) 'A point mutation in the HIV-1 Tat responsive element is associated with postintegration latency.', *Proceedings of the National Academy of Sciences of the United States of America*, 93(13), pp. 6377–81. doi: 10.1073/pnas.93.13.6377.
Erkelenz, S. *et al.* (2014) 'Genomic HEXploring allows landscaping of novel potential splicing regulatory elements', *Nucleic Acids Research*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 42(16), pp. 10681–10697. doi: 10.1093/nar/gku736.
Erkelenz, S. *et al.* (2015) 'Balanced splicing at the Tat-specific HIV-1 3'ss A3 is critical for HIV-1 replication.', *Retrovirology*. BioMed Central Ltd, 12(1), p. 29. doi: 10.1186/s12977-015-0154-8.
Falangola, M. F. *et al.* (1995) 'HIV infection of human choroid plexus: a possible mechanism of viral entry into the CNS.', *Journal of neuropathology and experimental neurology*, 54(4), pp. 497–503. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7602324 (Accessed: 12 January 2016).
Faria, N. R. *et al.* (2014) 'The early spread and epidemic ignition of HIV-1 in human populations', *Science*. American Association for the Advancement of Science, 346(6205), pp. 56–61. doi: 10.1126/science.1256739.
Fauci, A. S. and Marston, H. D. (2015) 'Toward an HIV vaccine: A scientific journey.', *Science (New York, N.Y.)*, 349(6246), pp. 386–7. doi: 10.1126/science.aac6300.
Feng, Q. *et al.* (2002) *Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain*, *Current Biology*. doi: 10.1016/S0960-9822(02)00901-6.
Fidler, S. *et al.* (2013) 'Short-course antiretroviral therapy in primary HIV infection.', *The New England journal of medicine*, 368, pp. 207–17. doi: 10.1056/NEJMoa1110039.
Finkelshtein, D. *et al.* (2013) 'LDL receptor and its family members serve as the

cellular receptors for vesicular stomatitis virus', *Proceedings of the National Academy of Sciences*, 110(18), pp. 7306–7311. doi: 10.1073/pnas.1214441110.
Finzi, D. *et al.* (1997) 'Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy.', *Science*, 278(5341), pp. 1295–1300.
Folks, T. *et al.* (1986) 'Induction of HTLV-III/LAV from a nonvirus-producing T-cell line: implications for latency', *Science*, 231(4738), pp. 600–602. doi: 10.1126/science.3003906.
Folks, T. M. *et al.* (1989) 'Tumor necrosis factor alpha induces expression of human immunodeficiency virus in a chronically infected T-cell clone.', *Proceedings of the National Academy of Sciences*, 86(7), pp. 2365–2368. doi: 10.1073/pnas.86.7.2365.
Fowke, K. R. *et al.* (1996) 'Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya.', *Lancet (London, England)*. Elsevier, 348(9038), pp. 1347–51. doi: 10.1016/S0140-6736(95)12269-2.
Friedman, J. *et al.* (2011) 'Epigenetic Silencing of HIV-1 by the Histone H3 Lysine 27 Methyltransferase Enhancer of Zeste 2', *Journal of Virology*. American Society for Microbiology, 85(17), pp. 9078–9089. doi: 10.1128/JVI.00836-11.
Fun, A. *et al.* (2017) 'A highly reproducible quantitative viral outgrowth assay for the measurement of the replication-competent latent HIV-1 reservoir.', *Scientific reports*. Nature Publishing Group, 7, p. 43231. doi: 10.1038/srep43231.
Gail, M. H. *et al.* (1997) 'Survival after AIDS diagnosis in a cohort of hemophilia patients. Multicenter Hemophilia Cohort Study.', *Journal of acquired immune deficiency syndromes and human retrovirology : official publication of the International Retrovirology Association*, 15(5), pp. 363–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9342256 (Accessed: 11 January 2016).
Gallastegui, E. *et al.* (2011) 'Chromatin reassembly factors are involved in transcriptional interference promoting HIV latency.', *Journal of virology*. American Society for Microbiology, 85(7), pp. 3187–202. doi: 10.1128/JVI.01920-10.
Gao, F. *et al.* (2004) 'Unselected mutations in the human immunodeficiency virus type 1 genome are mostly nonsynonymous and often deleterious.', *Journal of virology*. American Society for Microbiology, 78(5), pp. 2426–33. doi: 10.1128/JVI.78.5.2426-2433.2004.
Gifford, R. J. *et al.* (2007) 'Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1.', *Journal of virology*, 81(23), pp. 13050–6. doi: 10.1128/JVI.00889-07.
Goudsmit, J. *et al.* (1998) 'Genetic differences between human immunodeficiency virus type 1 subpopulations in faeces and serum.', *Journal of General Virology*, 79(2), pp. 259–267. doi: 10.1099/0022-1317-79-2-259.
Greenwood, E. J. *et al.* (2016) 'Temporal proteomic analysis of HIV infection reveals remodelling of the host phosphoproteome by lentiviral Vif variants', *eLife*. eLife Sciences Publications Limited, 5, p. e18296. doi: 10.7554/eLife.18296.
Grewal, S. I. S. and Jia, S. (2007) 'Heterochromatin revisited.', *Nature reviews. Genetics*. Nature Publishing Group, 8(1), pp. 35–46. doi: 10.1038/nrg2008.
Group, T. I. S. S. (2015) 'Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection', *New England Journal of Medicine*. Massachusetts Medical Society, 373(9), pp. 795–807. doi: 10.1056/NEJMoa1506816.
Gulick, R. M. *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(11), pp. 734–739. doi:

10.1056/NEJM199709113371102.
Hahn, S. (2004) 'Structure and mechanism of the RNA polymerase II transcription machinery', *Nature Structural & Molecular Biology*. Nature Publishing Group, 11(5), pp. 394–403. doi: 10.1038/nsmb763.
Hammer, S. M. *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(11), pp. 725–33. doi: 10.1056/NEJM199709113371101.
Han, Y. *et al.* (2004a) 'Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes.', *Journal of virology*, 78(12), pp. 6122–33. doi: 10.1128/JVI.78.12.6122-6133.2004.
Han, Y. *et al.* (2004b) 'Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes.', *Journal of virology*, 78(12), pp. 6122–33. Available at: http://jvi.asm.org/cgi/content/abstract/78/12/6122 (Accessed: 20 July 2012).
Haynes, B. F. *et al.* (2012) 'B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study.', *Nature biotechnology*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 30(5), pp. 423–33. doi: 10.1038/nbt.2197.
He, N. *et al.* (2010) 'HIV-1 Tat and Host AFF4 Recruit Two Transcription Elongation Factors into a Bifunctional Complex for Coordinated Activation of HIV-1 Transcription', *Molecular Cell*, 38(3), pp. 428–438. doi: 10.1016/j.molcel.2010.04.013.
Van Heuverswyn, F. *et al.* (2006) 'Human immunodeficiency viruses: SIV infection in wild gorillas.', *Nature*, 444(7116), p. 164. doi: 10.1038/444164a.
Ho, Y.-C. *et al.* (2013) 'Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure.', *Cell*. Elsevier, 155(3), pp. 540–51. doi: 10.1016/j.cell.2013.09.020.
Holman, A. G. and Coffin, J. M. (2005) 'Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites', *Proceedings of the National Academy of Sciences*, 102(17), pp. 6103–6107. doi: 10.1073/pnas.0501646102.
Hosmane, N. N. *et al.* (2017) 'Proliferation of latently infected CD4 + T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics', *The Journal of Experimental Medicine*, 214(4), pp. 959–972. doi: 10.1084/jem.20170193.
Hu, W.-S. and Hughes, S. H. (2012) 'HIV-1 reverse transcription.', *Cold Spring Harbor perspectives in medicine*. Cold Spring Harbor Laboratory Press, 2(10). doi: 10.1101/cshperspect.a006882.
Jenuwein, T. *et al.* (2000) 'Regulation of chromatin structure by site-specific histone H3 methyltransferases', *Nature*. Nature Publishing Group, 406(6796), pp. 593–599. doi: 10.1038/35020506.
Jones, P. A. (2012) 'Functions of DNA methylation: islands, start sites, gene bodies and beyond', *Nature Reviews Genetics*. Nature Publishing Group, 13(7), pp. 484–492. doi: 10.1038/nrg3230.
Jordan, A., Bisgrove, D. and Verdin, E. (2003) 'HIV reproducibly establishes a latent

infection after acute infection of T cells in vitro.', *The EMBO journal*. European Molecular Biology Organization, 22(8), pp. 1868–77. doi: 10.1093/emboj/cdg188.
Jordan, A., Defechereux, P. and Verdin, E. (2001) 'The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation.', *The EMBO journal*, 20(7), pp. 1726–38. doi: 10.1093/emboj/20.7.1726.
Kaminski, R. *et al.* (2016) 'Elimination of HIV-1 Genomes from Human T-lymphoid Cells by CRISPR/Cas9 Gene Editing.', *Scientific reports*. Nature Publishing Group, 6, p. 22555. doi: 10.1038/srep22555.
Karn, J. and Stoltzfus, C. M. (2012) 'Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression', *Cold Spring Harbor Perspectives in Medicine*, 2(2), pp. a006916–a006916. doi: 10.1101/cshperspect.a006916.
Kaul, R. *et al.* (2001) 'New insights into HIV-1 specific cytotoxic T-lymphocyte responses in exposed, persistently seronegative Kenyan sex workers', *Immunology Letters*. Elsevier, 79(1–2), pp. 3–13. doi: 10.1016/S0165-2478(01)00260-7.
Keele, B. F. *et al.* (2006) 'Chimpanzee reservoirs of pandemic and nonpandemic HIV-1.', *Science (New York, N.Y.)*, 313(5786), pp. 523–6. doi: 10.1126/science.1126531.
Keet, I. P. *et al.* (1994) 'Characteristics of long-term asymptomatic infection with human immunodeficiency virus type 1 in men with normal and low CD4+ cell counts.', *The Journal of infectious diseases*, 169(6), pp. 1236–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7910835 (Accessed: 26 February 2018).
Kessing, C. F. *et al.* (2017) 'In Vivo Suppression of HIV Rebound by Didehydro-Cortistatin A, a "Block-and-Lock" Strategy for HIV-1 Treatment.', *Cell reports*. Elsevier, 21(3), pp. 600–611. doi: 10.1016/j.celrep.2017.09.080.
Kireeva, M. L. *et al.* (2002) 'Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription.', *Molecular cell*, 9(3), pp. 541–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11931762 (Accessed: 12 May 2018).
Kokura, K. *et al.* (2010) 'Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion', *The EMBO Journal*, 29(21), pp. 3673–3687. doi: 10.1038/emboj.2010.239.
Komninakis, S. V *et al.* (2012) 'HIV-1 proviral DNA loads (as determined by quantitative PCR) in patients subjected to structured treatment interruption after antiretroviral therapy failure.', *Journal of clinical microbiology*. American Society for Microbiology, 50(6), pp. 2132–3. doi: 10.1128/JCM.00393-12.
Kornberg, R. D. (1974) 'Chromatin structure: a repeating unit of histones and DNA.', *Science (New York, N.Y.)*, 184(4139), pp. 868–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4825889 (Accessed: 23 February 2017).
Kornberg, R. D. and Thomas, J. O. (1974) 'Chromatin structure; oligomers of the histones.', *Science (New York, N.Y.)*, 184(4139), pp. 865–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4825888 (Accessed: 23 February 2017).
Kostrewa, D. *et al.* (2009) 'RNA polymerase II–TFIIB structure and mechanism of transcription initiation', *Nature*. Nature Publishing Group, 462(7271), pp. 323–330. doi: 10.1038/nature08548.
Kouzarides, T. (2007) 'Chromatin modifications and their function.', *Cell*. Elsevier, 128(4), pp. 693–705. doi: 10.1016/j.cell.2007.02.005.
Krogan, N. J. *et al.* (2003) 'The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone

methylation.', *Molecular cell*. Elsevier, 11(3), pp. 721–9. doi: 10.1016/S1097-2765(03)00091-1.
Krueger, B. J. *et al.* (2010) 'The mechanism of release of P-TEFb and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK.', *PloS one*. Edited by M. V. Blagosklonny. Public Library of Science, 5(8), p. e12335. doi: 10.1371/journal.pone.0012335.
Kurita, M. *et al.* (2007) 'CR/periphilin is a transcriptional co-repressor involved in cell cycle progression', *Biochemical and Biophysical Research Communications*, 364(4), pp. 930–936. doi: 10.1016/j.bbrc.2007.10.090.
Kuzmichev, Y. V *et al.* (2017) 'A CD3/CD28 microbead-based HIV-1 viral outgrowth assay.', *Journal of virus eradication*. Mediscript, 3(2), pp. 85–89. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28435692 (Accessed: 20 July 2017).
Laird, G. M. *et al.* (2013) 'Rapid Quantification of the Latent Reservoir for HIV-1 Using a Viral Outgrowth Assay', *PLoS Pathogens*. Edited by G. Silvestri. Public Library of Science, 9(5), p. e1003398. doi: 10.1371/journal.ppat.1003398.
Lassen, K. G. *et al.* (2012) 'A flexible model of HIV-1 latency permitting evaluation of many primary CD4 T-cell reservoirs.', *PloS one*. Edited by R. F. Speck. Public Library of Science, 7(1), p. e30176. doi: 10.1371/journal.pone.0030176.
Lee, G. Q. *et al.* (2017) 'Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells', *Journal of Clinical Investigation*. American Society for Clinical Investigation. doi: 10.1172/JCI93289.
Li, J. Z. *et al.* (2014) 'Relationship of HIV reservoir characteristics with immune status and viral rebound kinetics in an HIV therapeutic vaccine study.', *AIDS (London, England)*. NIH Public Access, 28(18), pp. 2649–57. doi: 10.1097/QAD.0000000000000478.
Liao, H. K. *et al.* (2015) 'Use of the CRISPR/Cas9 system as an intracellular defense against HIV-1 infection in human cells', *Nature Communications*. Nature Publishing Group, 6, p. 6413. doi: 10.1038/ncomms7413.
Liao, J. *et al.* (2017) 'Characterization of retroviral infectivity and superinfection resistance during retrovirus-mediated transduction of mammalian cells', *Gene Therapy*. Nature Publishing Group, 24(6), pp. 333–341. doi: 10.1038/gt.2017.24.
Limsirichai, P., Gaj, T. and Schaffer, D. V (2016) 'CRISPR-mediated Activation of Latent HIV-1 Expression', *Molecular Therapy*. Cell Press, 24(3), pp. 499–507. doi: 10.1038/MT.2015.213.
Van Lint, C., Bouchat, S. and Marcello, A. (2013) 'HIV-1 transcription and latency: an update.', *Retrovirology*, 10(1), p. 67. doi: 10.1186/1742-4690-10-67.
Liu, L. *et al.* (2011) 'A whole genome screen for HIV restriction factors', *Retrovirology*. BioMed Central, 8(1), p. 94. doi: 10.1186/1742-4690-8-94.
Lorch, Y., LaPointe, J. W. and Kornberg, R. D. (1987) 'Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones', *Cell*, 49(2), pp. 203–210. doi: 10.1016/0092-8674(87)90561-7.
Lorenzo-Redondo, R. *et al.* (2016) 'Persistent HIV-1 replication maintains the tissue reservoir during therapy', *Nature*. Nature Publishing Group, 530(7588), pp. 51–56. Available at: http://dx.doi.org/10.1038/nature16933%5Cn10.1038/nature16933%5Cnhttp://www.nature.com/nature/journal/v530/n7588/abs/nature16933.html#supplementary-information%5Cnhttp://www.nature.com/doifinder/10.1038/nature16933.

Luger, K. *et al.* (1997) 'Crystal structure of the nucleosome core particle at 2.8 A resolution.', *Nature*, 389(6648), pp. 251–60. doi: 10.1038/38444.
Lusic, M. *et al.* (2003) 'Regulation of HIV-1 gene expression by histone acetylation and factor recruitment at the LTR promoter.', *The EMBO journal*, 22(24), pp. 6550–61. doi: 10.1093/emboj/cdg631.
Luzuriaga, K. *et al.* (2014) 'HIV type 1 (HIV-1) proviral reservoirs decay continuously under sustained virologic control in HIV-1-infected children who received early treatment.', *The Journal of infectious diseases*, 210(10), pp. 1529–38. doi: 10.1093/infdis/jiu297.
Maldarelli, F. *et al.* (2014) 'Specific HIV integration sites are linked to clonal expansion and persistence of infected cells', *Science*, 345(6193), pp. 179–183. doi: 10.1126/science.1254194.
Mansky, L. M. and Temin, H. M. (1995) 'Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase.', *Journal of virology*. American Society for Microbiology, 69(8), pp. 5087–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7541846 (Accessed: 27 February 2018).
Marini, B. *et al.* (2015) 'Nuclear architecture dictates HIV-1 integration site selection', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. doi: 10.1038/nature14226.
Martin Stoltzfus, C. (2009) 'Chapter 1 Regulation of HIV-1 Alternative RNA Splicing and Its Role in Virus Replication', *Advances in Virus Research*, pp. 1–40. doi: 10.1016/S0065-3527(09)74001-1.
Martínez-Bonet, M. *et al.* (2015) 'Synergistic Activation of Latent HIV-1 Expression by Novel Histone Deacetylase Inhibitors and Bryostatin-1', *Scientific Reports*. Nature Publishing Group, 5(1), p. 16445. doi: 10.1038/srep16445.
Marx, J. L. (1985) 'A virus by any other name . . .', *Science (New York, N.Y.)*, 227(4693), pp. 1449–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2983427 (Accessed: 7 December 2015).
Marzio, G. *et al.* (1998) 'HIV-1 tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter.', *Proceedings of the National Academy of Sciences of the United States of America*, 95(23), pp. 13519–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9811832 (Accessed: 28 February 2018).
Matheson, N. J., Peden, A. A. and Lehner, P. J. (2014) 'Antibody-free magnetic cell sorting of genetically modified primary human CD4+ T cells by one-step streptavidin affinity purification', *PLoS ONE*. Edited by T. A. Baldwin, 9(10), p. e111437. doi: 10.1371/journal.pone.0111437.
Matsui, T. *et al.* (1980) 'Multiple factors required for accurate initiation of transcription by purified RNA polymerase II.', *Journal of Biological Chemistry* , 255(24), pp. 11992–11996. Available at: http://www.jbc.org/content/255/24/11992.abstract.
Mehla, R. *et al.* (2010) 'Bryostatin modulates latent HIV-1 infection via PKC and AMPK signaling but inhibits acute infection in a receptor independent manner.', *PloS one*. Public Library of Science, 5(6), p. e11160. doi: 10.1371/journal.pone.0011160.
Miyazato, P. *et al.* (2016) 'Transcriptional and Epigenetic Regulatory Mechanisms Affecting HTLV-1 Provirus.', *Viruses*. Multidisciplinary Digital Publishing Institute (MDPI), 8(6). doi: 10.3390/v8060171.

MMWR (1982a) 'A cluster of Kaposi's sarcoma and Pneumocystis carinii pneumonia among homosexual male residents of Los Angeles and Orange Counties, California.', *MMWR. Morbidity and mortality weekly report*, 31(23), pp. 305–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6811844 (Accessed: 7 December 2015).
MMWR (1982b) 'Persistent, generalized lymphadenopathy among homosexual males.', *MMWR. Morbidity and mortality weekly report*, 31(19), pp. 249–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6808340 (Accessed: 7 December 2015).
MMWR (1982c) 'Pneumocystis carinii pneumonia among persons with hemophilia A.', *MMWR. Morbidity and mortality weekly report*, 31(27), pp. 365–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6815443 (Accessed: 7 December 2015).
Mok, H. P. *et al.* (2018) 'No evidence of ongoing evolution in replication competent latent HIV-1 in a patient followed up for two years', *Scientific Reports 2018 8:1*. Nature Publishing Group, 8(1), p. 2639. doi: 10.1038/s41598-018-20682-w.
Moore, C. B. *et al.* (2010) 'Short hairpin RNA (shRNA): design, delivery, and assessment of gene knockdown.', *Methods in molecular biology (Clifton, N.J.)*. NIH Public Access, 629, pp. 141–58. doi: 10.1007/978-1-60761-657-3_10.
Myers, L. E., McQuay, L. J. and Hollinger, F. B. (1994) 'Dilution assay statistics.', *Journal of clinical microbiology*. American Society for Microbiology, 32(3), pp. 732–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8195386 (Accessed: 20 July 2017).
Nelson, T. J. *et al.* (2017) 'Bryostatin Effects on Cognitive Function and PKCε in Alzheimer's Disease Phase IIa and Expanded Access Trials.', *Journal of Alzheimer's disease : JAD*. IOS Press, 58(2), pp. 521–535. doi: 10.3233/JAD-170161.
Nethe, M., Berkhout, B. and van der Kuyl, A. C. (2005) 'Retroviral superinfection resistance', *Retrovirology*. BioMed Central, 2(1), p. 52. doi: 10.1186/1742-4690-2-52.
Ngo-Giang-Huong, N. *et al.* (2001) 'Proviral HIV-1 DNA in subjects followed since primary HIV-1 infection who suppress plasma viral load after one year of highly active antiretroviral therapy.', *AIDS (London, England)*, 15(6), pp. 665–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11371680 (Accessed: 13 January 2015).
Nguyen, K. *et al.* (2017) 'Multiple histone lysine methyltransferases are required for the establishment and maintenance of HIV-1 latency', *mBio*. American Society for Microbiology (ASM), 8(1). doi: 10.1128/mBio.00133-17.
Noë, A., Plum, J. and Verhofstede, C. (2005) 'The latent HIV-1 reservoir in patients undergoing HAART: an archive of pre-HAART drug resistance.', *The Journal of antimicrobial chemotherapy*, 55(4), pp. 410–2. doi: 10.1093/jac/dki038.
O'Brien, S. K. *et al.* (2010) 'P-TEFb kinase complex phosphorylates histone H1 to regulate expression of cellular and HIV-1 genes.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 285(39), pp. 29713–20. doi: 10.1074/jbc.M110.125997.
Ocwieja, K. E. *et al.* (2012) 'Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing', *Nucleic Acids Research*. Oxford University Press, 40(20), pp. 10345–10355. doi: 10.1093/nar/gks753.
Oliveira, M. F. *et al.* (2017) 'Early Antiretroviral Therapy Is Associated with Lower HIV DNA Molecular Diversity and Lower Inflammation in Cerebrospinal Fluid but

Does Not Prevent the Establishment of Compartmentalized HIV DNA Populations', *PLOS Pathogens*. Edited by D. C. Douek. Public Library of Science, 13(1), p. e1006112. doi: 10.1371/journal.ppat.1006112.
Ostrowski, M. A. *et al.* (1999) 'Both memory and CD45RA+/CD62L+ naive CD4(+) T cells are infected in human immunodeficiency virus type 1-infected individuals.', *Journal of virology*. American Society for Microbiology, 73(8), pp. 6430–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10400736 (Accessed: 30 April 2018).
Ott, M. *et al.* (1999) 'Acetylation of the HIV-1 tat protein by p300 is important for its transcriptional activity', *Current Biology*, 9(24), pp. 1489–1492. doi: 10.1016/S0960-9822(00)80120-7.
Passaes, C. P. B. *et al.* (2017) 'Ultrasensitive HIV-1 p24 Assay Detects Single Infected Cells and Differences in Reservoir Induction by Latency Reversal Agents.', *Journal of virology*. American Society for Microbiology, 91(6), pp. e02296-16. doi: 10.1128/JVI.02296-16.
Pearson, R. *et al.* (2008) 'Epigenetic silencing of human immunodeficiency virus (HIV) transcription by formation of restrictive chromatin structures at the viral long terminal repeat drives the progressive entry of HIV into latency.', *Journal of virology*, 82(24), pp. 12291–303. doi: 10.1128/JVI.01383-08.
Persaud, D. *et al.* (2013) 'Absence of detectable HIV-1 viremia after treatment cessation in an infant.', *The New England journal of medicine*, 369(19), pp. 1828–35. doi: 10.1056/NEJMoa1302976.
Philpott, S. *et al.* (1999) 'CCR5 genotype and resistance to vertical transmission of HIV-1.', *Journal of acquired immune deficiency syndromes (1999)*, 21(3), pp. 189–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10421241 (Accessed: 11 January 2016).
Preston, B. D., Poiesz, B. J. and Loeb, L. A. (1988) 'Fidelity of HIV-1 reverse transcriptase.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 242(4882), pp. 1168–71. doi: 10.1126/SCIENCE.2460924.
Procopio, F. A. *et al.* (2015) 'A Novel Assay to Measure the Magnitude of the Inducible Viral Reservoir in HIV-infected Individuals', *EBioMedicine*, 2(8). doi: 10.1016/j.ebiom.2015.06.019.
Rasmussen, T. A. *et al.* (2013) 'Comparison of HDAC inhibitors in clinical development: effect on HIV production in latently infected cells and T-cell activation.', *Human vaccines & immunotherapeutics*. Landes Bioscience, 9(5), pp. 993–1001. doi: 10.4161/hv.23800.
Rasmussen, T. A. *et al.* (2014) 'Panobinostat, a histone deacetylase inhibitor, for latent-virus reactivation in HIV-infected patients on suppressive antiretroviral therapy: a phase 1/2, single group, clinical trial', *The Lancet HIV*. Elsevier, 1(1), pp. e13–e21. doi: 10.1016/S2352-3018(14)70014-1.
Revyakin, A. *et al.* (2006) 'Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching.', *Science (New York, N.Y.)*. NIH Public Access, 314(5802), pp. 1139–43. doi: 10.1126/science.1131398.
Rissin, D. M. *et al.* (2010) 'Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations', *Nature Biotechnology*, 28(6), pp. 595–599. doi: 10.1038/nbt.1641.
Rivière, L. *et al.* (2015) 'HBx relieves chromatin-mediated transcriptional repression of hepatitis B viral cccDNA involving SETDB1 histone

methyltransferase', *Journal of Hepatology*, 63(5), pp. 1093–1102. doi: 10.1016/j.jhep.2015.06.023.
Roberts, J. D., Bebenek, K. and Kunkel, T. A. (1988) 'The accuracy of reverse transcriptase from HIV-1.', *Science (New York, N.Y.)*, 242(4882), pp. 1171–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2460925 (Accessed: 10 December 2017).
Rosonina, E. and Manley, J. L. (2005) 'From Transcription to mRNA: PAF Provides a New Path', *Molecular Cell*. Cell Press, 20(2), pp. 167–168. doi: 10.1016/J.MOLCEL.2005.10.004.
Rouzine, I. M., Weinberger, A. D. and Weinberger, L. S. (2015) 'An Evolutionary Role for HIV Latency in Enhancing Viral Transmission', *Cell*, 160(5), pp. 1002–1012. doi: 10.1016/j.cell.2015.02.017.
Rowland-Jones, S. *et al.* (1995) 'HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women', *Nature Medicine*. Nature Publishing Group, 1(1), pp. 59–64. doi: 10.1038/nm0195-59.
Saag, M. and Deeks, S. G. (2010) 'How Do HIV Elite Controllers Do What They Do?', *Clinical Infectious Diseases*. Oxford University Press, 51(2), pp. 239–241. doi: 10.1086/653678.
Sáez-Cirión, A. *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*. Edited by J. Lifson. Public Library of Science, 9(3), p. e1003211. doi: 10.1371/journal.ppat.1003211.
Salazar-Gonzalez, J. F. *et al.* (2009) 'Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection.', *The Journal of experimental medicine*. The Rockefeller University Press, 206(6), pp. 1273–89. doi: 10.1084/jem.20090378.
Saleh, S. *et al.* (2007) 'CCR7 ligands CCL19 and CCL21 increase permissiveness of resting memory CD4+ T cells to HIV-1 infection: a novel model of HIV-1 latency.', *Blood*, 110(13), pp. 4161–4. doi: 10.1182/blood-2007-06-097907.
Saleh, S. *et al.* (2016) 'HIV integration and the establishment of latency in CCL19-treated resting CD4+ T cells require activation of NF-κB', *Retrovirology*. BioMed Central, 13(1), p. 49. doi: 10.1186/s12977-016-0284-7.
Sander, J. D. and Joung, J. K. (2014) 'CRISPR-Cas systems for editing, regulating and targeting genomes.', *Nature biotechnology*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 32(4), pp. 347–55. doi: 10.1038/nbt.2842.
Sanyal, A. *et al.* (2017) 'Novel assay reveals a large, inducible, replication-competent HIV-1 reservoir in resting CD4+ T cells', *Nature Medicine*. Nature Research, 23(7), pp. 885–889. doi: 10.1038/nm.4347.
Schneider, U., Schwenk, H. U. and Bornkamm, G. (1977) 'Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma.', *International journal of cancer*, 19(5), pp. 621–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/68013 (Accessed: 10 June 2018).
Schröder, A. R. W. *et al.* (2002) 'HIV-1 integration in the human genome favors active genes and local hotspots.', *Cell*, 110(4), pp. 521–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12202041 (Accessed: 5 January 2016).
Schwabish, M. A. and Struhl, K. (2004) 'Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II.', *Molecular and*

cellular biology. American Society for Microbiology, 24(23), pp. 10111–7. doi: 10.1128/MCB.24.23.10111-10117.2004.
Shan, L. *et al.* (2012) 'Stimulation of HIV-1-specific cytolytic T lymphocytes facilitates elimination of latent viral reservoir after virus reactivation.', *Immunity*, 36(3), pp. 491–501. doi: 10.1016/j.immuni.2012.01.014.
Sharp, P. M. *et al.* (2001) 'The origins of acquired immune deficiency syndrome viruses: where and when?', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 356(1410), pp. 867–76. doi: 10.1098/rstb.2001.0863.
Sharp, P. M. and Hahn, B. H. (2011) 'Origins of HIV and the AIDS pandemic.', *Cold Spring Harbor perspectives in medicine*, 1(1), p. a006841. doi: 10.1101/cshperspect.a006841.
Sheppard, H. W. *et al.* (1993) 'The characterization of non-progressors: Long-term HIV-1 infection with stable CD4+ T-cell levels', *AIDS*, 7(9), pp. 1159–1166. doi: 10.1097/00002030-199309000-00002.
Shingai, M. *et al.* (2014) 'Passive transfer of modest titers of potent and broadly neutralizing anti-HIV monoclonal antibodies block SHIV infection in macaques.', *The Journal of experimental medicine*, 211(10), pp. 2061–74. doi: 10.1084/jem.20132494.
Siliciano, J. M. and Siliciano, R. F. (2015) 'The Remarkable Stability of the Latent Reservoir for HIV-1 in Resting Memory CD4+ T Cells.', *The Journal of infectious diseases*, p. jiv219-. doi: 10.1093/infdis/jiv219.
Simic, R. *et al.* (2003) 'Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes', *The EMBO Journal*, 22(8), pp. 1846–1856. doi: 10.1093/emboj/cdg179.
Simonetti, F. R. *et al.* (2016) 'Clonally expanded CD4+ T cells can produce infectious HIV-1 in vivo.', *Proceedings of the National Academy of Sciences of the United States of America*, 113(7), pp. 1883–1888. doi: 10.1073/pnas.1522675113.
Singh, K. K. and Spector, S. A. (2009) 'Host genetic determinants of human immunodeficiency virus infection and disease progression in children.', *Pediatric research*. NIH Public Access, 65(5 Pt 2), p. 55R–63R. doi: 10.1203/PDR.0b013e31819dca03.
Smith, M. W. *et al.* (1997) 'Contrasting Genetic Influence of CCR2 and CCR5 Variants on HIV-1 Infection and Disease Progression', *Science*. American Association for the Advancement of Science, 277(5328), pp. 959–965. doi: 10.1126/science.277.5328.959.
Spina, C. A. *et al.* (2013) 'An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients.', *PLoS pathogens*. Edited by M. Emerman. Public Library of Science, 9(12), p. e1003834. doi: 10.1371/journal.ppat.1003834.
Spina, C. A., Guatelli, J. C. and Richman, D. D. (1995) 'Establishment of a stable, inducible form of human immunodeficiency virus type 1 DNA in quiescent CD4 lymphocytes in vitro.', *Journal of virology*, 69(5), pp. 2977–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7707524 (Accessed: 26 March 2018).
Stevenson, M. *et al.* (1990) 'HIV-1 replication is controlled at the level of T cell activation and proviral integration.', *The EMBO journal*, 9(5), pp. 1551–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2184033 (Accessed: 1 March 2018).
Stopak, K. *et al.* (2003) 'HIV-1 Vif Blocks the Antiviral Activity of APOBEC3G by

Impairing Both Its Translation and Intracellular Stability', *Molecular Cell*, 12(3), pp. 591–601. doi: 10.1016/S1097-2765(03)00353-8.
Strahl, B. D. and Allis, C. D. (2000) 'The language of covalent histone modifications', *Nature*. Nature Publishing Group, 403(6765), pp. 41–45. doi: 10.1038/47412.
Struhl, K. (1998) 'Histone acetylation and transcriptional regulatory mechanisms.', *Genes & development*. Cold Spring Harbor Laboratory Press, 12(5), pp. 599–606. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9499396 (Accessed: 12 May 2018).
Swiggard, W. J. *et al.* (2005) 'Human immunodeficiency virus type 1 can establish latent infection in resting CD4+ T cells in the absence of activating stimuli.', *Journal of virology*. American Society for Microbiology, 79(22), pp. 14179–88. doi: 10.1128/JVI.79.22.14179-14188.2005.
Symons, J. *et al.* (2017) 'HIV integration sites in latently infected cell lines: evidence of ongoing replication.', *Retrovirology*. BioMed Central, 14(1), p. 2. doi: 10.1186/s12977-016-0325-2.
Tchasovnikarova, I. A. *et al.* (2015) 'Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells', *Science*. American Association for the Advancement of Science, 348(6242), pp. 1481–1485. doi: 10.1126/science.aaa7227.
Thierry, S. *et al.* (2015) 'Integrase inhibitor reversal dynamics indicate unintegrated HIV-1 dna initiate de novo integration.', *Retrovirology*, 12(1), p. 24. doi: 10.1186/s12977-015-0153-9.
Townsend, C. L. *et al.* (2008) 'Low rates of mother-to-child transmission of HIV following effective pregnancy interventions in the United Kingdom and Ireland, 2000-2006.', *AIDS (London, England)*, 22(8), pp. 973–81. doi: 10.1097/QAD.0b013e3282f9b67a.
Tyagi, M., Pearson, R. J. and Karn, J. (2010) 'Establishment of HIV latency in primary CD4+ cells is due to epigenetic transcriptional silencing and P-TEFb restriction.', *Journal of virology*, 84(13), pp. 6425–37. doi: 10.1128/JVI.01519-09.
Utley, R. T. *et al.* (1997) 'SWI/SNF stimulates the formation of disparate activator-nucleosome complexes but is partially redundant with cooperative binding.', *The Journal of biological chemistry*, 272(19), pp. 12642–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9139720 (Accessed: 12 May 2018).
Verdin, E., Paras, P. and Van Lint, C. (1993) 'Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation.', *The EMBO journal*, 12(8), pp. 3249–59. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=413592&tool=pmcentrez&rendertype=abstract (Accessed: 9 December 2015).
Wagner, T. A. *et al.* (2013) 'An increasing proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment suggests proliferation of HIV-infected cells.', *Journal of virology*, 87(3), pp. 1770–8. doi: 10.1128/JVI.01985-12.
Wagner, T. A. *et al.* (2014) 'Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection', *Science*, 345(6196). doi: 10.1126/science.1256304.
Wang, G. P. *et al.* (2007) 'HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications.', *Genome research*. Cold Spring Harbor Laboratory Press, 17(8), pp. 1186–94. doi: 10.1101/gr.6286907.
Weinberger, L. S. *et al.* (2005) 'Stochastic gene expression in a lentiviral positive-

feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity.', *Cell*, 122(2), pp. 169–82. doi: 10.1016/j.cell.2005.06.006.
West, M. J. and Karn, J. (1999) 'Stimulation of Tat-associated kinase-independent transcriptional elongation from the human immunodeficiency virus type-1 long terminal repeat by a cellular enhancer', *The EMBO Journal*, 18(5), p. 1378 LP-1386. Available at: http://emboj.embopress.org/content/18/5/1378.abstract.
WHO (2015) 'WHO | HIV/AIDS'. World Health Organization. Available at: http://www.who.int/mediacentre/factsheets/fs360/en/ (Accessed: 7 December 2015).
Wilkinson, D. A. *et al.* (1998) 'A 32-bp Deletion within the CCR5 Locus Protects against Transmission of Parenterally Acquired Human Immunodeficiency Virus but Does Not Affect Progression to AIDS-Defining Illness', *The Journal of Infectious Diseases*, 178(4), pp. 1163–1166. doi: 10.1086/515675.
Winckelmann, A. *et al.* (2017) 'Romidepsin-induced HIV-1 viremia during effective antiretroviral therapy contains identical viral sequences with few deleterious mutations', *AIDS*, 31(6), pp. 771–779. doi: 10.1097/QAD.0000000000001400.
Winkler, D. D. and Luger, K. (2011) 'The histone chaperone FACT: structural insights and mechanisms for nucleosome reorganization.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 286(21), pp. 18369–74. doi: 10.1074/jbc.R110.180778.
Wong, J. *et al.* (1997) 'Recovery of replication-competent HIV despite prolonged suppression of plasma viremia', *Science*, 278(5341), pp. 1291–1295.
Wood, A. *et al.* (2003) 'The Paf1 Complex Is Essential for Histone Monoubiquitination by the Rad6-Bre1 Complex, Which Signals for Histone Methylation by COMPASS and Dot1p', *Journal of Biological Chemistry*, 278(37), pp. 34739–34742. doi: 10.1074/jbc.C300269200.
Workman, J. L. and Kingston, R. E. (1998) 'Alteration of nucleosome structure as a mechanism of trascriptional regulation', *Annual Review of Biochemistry*, 67(1), pp. 545–579. doi: 10.1146/annurev.biochem.67.1.545.
Worobey, M. *et al.* (2008) 'Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960.', *Nature*, 455(7213), pp. 661–4. doi: 10.1038/nature07390.
Wu, X. *et al.* (2005) 'Weak Palindromic Consensus Sequences Are a Common Feature Found at the Integration Target Sites of Many Retroviruses', *Journal of Virology*, 79(8), pp. 5211–5214. doi: 10.1128/JVI.79.8.5211-5214.2005.
Xing, S. *et al.* (2011) 'Disulfiram reactivates latent HIV-1 in a Bcl-2-transduced primary CD4+ T cell model without inducing global T cell activation.', *Journal of virology*, 85(12), pp. 6060–4. doi: 10.1128/JVI.02033-10.
Yamaguchi, Y. *et al.* (1999) 'NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation', *Cell*, 97(1), pp. 41–51. doi: 10.1016/S0092-8674(00)80713-8.
Yang, H.-C. *et al.* (2009) 'Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation', *Journal of Clinical Investigation*, 119(11), pp. 3473–86. doi: 10.1172/JCI39199.
Yerly, S. *et al.* (2004) 'Proviral HIV-DNA predicts viral rebound and viral setpoint after structured treatment interruptions.', *AIDS (London, England)*, 18(14), pp. 1951–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15353981 (Accessed: 2 April 2018).
Zack, J. A. *et al.* (1992) 'Incompletely reverse-transcribed human

immunodeficiency virus type 1 genomes in quiescent cells can function as intermediates in the retroviral life cycle.', *Journal of virology*, 66(3), pp. 1717–25. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=240919&tool=pmcentrez&rendertype=abstract (Accessed: 8 December 2015).
Zhang, X. H. F. and Chasin, L. A. (2004) 'Computational definition of sequence motifs governing constitutive exon splicing', *Genes and Development*. Cold Spring Harbor Laboratory Press, 18(11), pp. 1241–1250. doi: 10.1101/gad.1195304.
Zhou, Y. *et al.* (2005) 'Kinetics of human immunodeficiency virus type 1 decay following entry into resting CD4+ T cells.', *Journal of virology*, 79(4), pp. 2199–210. doi: 10.1128/JVI.79.4.2199-2210.2005.