Spring 2010

Lentiviral Integration Site Targeting: Host Determinants and Consequences

Keshet Ronen
University of Pennsylvania, keshet@upenn.edu

Lentiviral Integration Site Targeting: Host Determinants and Consequences

Abstract

A necessary step in the retroviral lifecycle is integration, the covalent insertion of the viral cDNA into the genome of the infected cell. This means that retroviruses, for example HIV, establish life-long infection. It also means that retroviruses are used as gene-delivery vectors to treat genetic diseases. Integration events are distributed non-randomly in the genome of the infected cell, with characteristic genus-specific preferences. In this dissertation, we focus on the lentiviral class of retroviruses, and explore two aspects of their integration: the means by which integration is targeted to its favored sites, and the consequences of integration at these sites for the host cell. The host protein LEDGF/p75 has been shown to interact with lentiviral integrases and contribute to their preference for integration in genes. We sought to establish the extent to which integration site selection is determined by LEDGF/p75 tethering. We first asked whether LEDGF/p75 was an essential integration tether, by analyzing integration site distribution in cells stringently depleted for LEDGF/p75. We found that LEDGF/p75 is responsible for much of the lentiviral integration preference, though probably not all. Secondly, we asked whether LEDGF/p75 tethering is sufficient to determine the genomic distribution of lentiviral integration. We used a fusion of LEDGF/p75's integrase-binding domain and the heterochromatin-binding protein CBX1 to show that lentiviral integration could be retargeted away from its usual distribution and into CBX1-bound regions. These results underline LEDGF/p75's central role in lentiviral integration, and the potential for manipulating its interaction with integrase. The effect of retroviral integration on the host cell is of particular relevance in gene therapy, where insertional activation of proto-oncogenes in patients is a serious concern. We present data on the genomic integration site distribution of a lentiviral vector for the correction of β-thalassemia in mice. While use of the same vector in a human patient led to clonal outgrowth, we report no evidence of insertional activation in the mouse model, but instead the suggestion that integration in genes may impart a growth disadvantage. This argues for the safety of lentiviral vectors, but raises questions about their effect on host gene expression.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Cell & Molecular Biology

First Advisor

Frederic Bushman

Keywords

HIV, lentivirus, integration, LEDGF, retrovirus, retroviral

Subject Categories

Virology

LENTIVIRAL INTEGRATION SITE TARGETING: HOST DETERMINANTS AND CONSEQUENCES

Keshet Ronen

A DISSERTATION

In

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

In Partial Fulfillment of the Requirements of the Degree of Doctor of Philosophy

2010

_____

Frederic D. Bushman, PhD.

Supervisor of Dissertation

_____

Daniel S. Kessler, PhD.

Graduate Group Chair

Dissertation Committee:

Paul Bates, PhD.

Shelley Berger, PhD.

James Hoxie, MD.

Haig Kazazian, MD.

DEDICATION

To my parents, with gratitude for their unwavering support and faith in my abilities, and for the opportunities they gave me access to.

ACKNOWLEDGEMENTS

I thank my advisor, Rick Bushman, for his ideas, guidance and energy during the completion of this research, and, of course, for providing the resources with which it was completed. I also thank my thesis committee, Paul Bates, Shelley Berger, Jim Hoxie, and Haig Kazazian, for their advice and encouragement. I am greatly indebted to fellow members of the Bushman lab for their central roles in my training as a scientist. In particular I am grateful to Angela Ciuffi, who taught me the procedures of retroviral integration site amplification and whose notebook I still refer to, and to Tracy Diamond and Christian Hoffmann, for endless discussions, suggestions and support. I aspire to one day have the skill, confidence and congeniality of these three colleagues. I thank Nirav Malani for his patient discussions and careful computational analysis, Gary Wang for advice on integration site analysis and Young Hwang for sharing his technical wisdom.

I am especially indebted to Bob Doms and Jim Hoxie, outstandingly generous mentors whose attention to my personal and professional development was enormously appreciated.

Completion of this dissertation would not have been possible without the support and perspective of my family and friends, motivating and soothing me during the years of experiments and the final months of writing. I am especially grateful to the Ronen clan (Sabrina, Shuki, Shelly, Tibien and Ziggy), Kiera Moynihan, Sarah Zlotnik, Thomas Ward, Jeremy Pine, Crystal Biruk, Amy Paeth, Kunal Patel,

Amanda Tuffli and Augustine Harawa. I thank all members of the Bushman lab for their friendship, understanding, commiseration and laughter over the last few years: Tracy Diamond, Nana Minkah, Matt Culyba, Gary Wang, Anna Kline, Mali Skotheim, Karen Ocwieja, Jenn Hwang, Nirav Malani, Rebecca Custers-Allen, Kyle Bittinger and especially Christian Hoffmann, bay-mate extraordinaire.

ABSTRACT

LENTIVIRAL INTEGRATION SITE TARGETING: HOST DETERMINANTS AND CONSEQUENCES

Keshet Ronen

Advisor: Frederic D. Bushman

A necessary step in the retroviral lifecycle is integration, the covalent insertion of the viral cDNA into the genome of the infected cell. This means that retroviruses, for example HIV, establish life-long infection. It also means that retroviruses are used as gene-delivery vectors to treat genetic diseases. Integration events are distributed non-randomly in the genome of the infected cell, with characteristic genus-specific preferences. This dissertation focuses on the lentiviral class of retroviruses, and explores two aspects of their integration: the means by which integration is targeted to its favored sites, and the consequences of integration at these sites for the host cell. The host protein LEDGF/p75 has been shown to interact with lentiviral integrases and contribute to their preference for integration in genes. We sought to establish the extent to which integration site selection is determined by LEDGF/p75 tethering. We first asked whether LEDGF/p75 was an essential integration tether, by analyzing integration site distribution in cells stringently depleted for LEDGF/p75. We found that LEDGF/p75 is responsible for much of the lentiviral integration preference, though probably not all. Secondly, we asked

whether LEDGF/p75 tethering is sufficient to determine the genomic distribution of lentiviral integration. We used a fusion of LEDGF/p75's integrase-binding domain and the heterochromatin-binding protein CBX1 to show that lentiviral integration could be retargeted away from its usual distribution and into CBX1-bound regions. These results underline LEDGF/p75's central role in lentiviral integration, and the potential for manipulating its interaction with integrase. The effect of retroviral integration on the host cell is of particular relevance in gene therapy, where insertional activation of proto-oncogenes in patients is a serious concern. We present data on the genomic integration site distribution of a lentiviral vector for the correction of β-thalassemia in mice. While use of the same vector in a human patient led to clonal outgrowth, we report no evidence of insertional activation in the mouse model, but instead the suggestion that integration in genes may impart a growth disadvantage. This argues for the safety of lentiviral vectors, but raises questions about their effect on host gene expression.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

TABLES

FIGURES

CHAPTER 1 – INTRODUCTION

1.1 Properties and lifecycle of the *Retroviridae*

Retroviruses are a diverse class of spherical, enveloped RNA viruses belonging to the family *Retroviridae*. Retroviral virions are 80-150nm in diameter and composed of two copies of a single-stranded positive-sense RNA genome, surrounded by a protein core, enveloped in a lipid bilayer of host cell origin studded with viral glycoproteins. Two steps in the viral lifecycle define the group: reverse transcription of the single-stranded RNA genome into double-stranded DNA, and subsequent integration of this DNA copy of the genome into the genome of the infected cell.

A schematic of the virion structure and lifecycle of retroviruses is illustrated in Figure 1-1. Following interaction of viral envelope proteins with a cellular receptor, membrane fusion takes place, either at the surface of the host cell or after endocytosis, depending on the retrovirus [1-4] and the viral core is released into the cytoplasm [5]. The core is made up of, in addition to the RNA genome dimer, the structural proteins matrix (MA), capsid (CA) and nucleocapsid (NC), and the replication enzymes protease (PR), reverse transcriptase (RT) and integrase (IN). Accessory proteins may also be present depending on the virus, as well as other molecules from the host cell [6-8]. After entry, the core undergoes a poorly characterized change in core composition known as uncoating [9]. Reverse transcription, mediated by virion-packaged RT, takes place [10] and the core moves

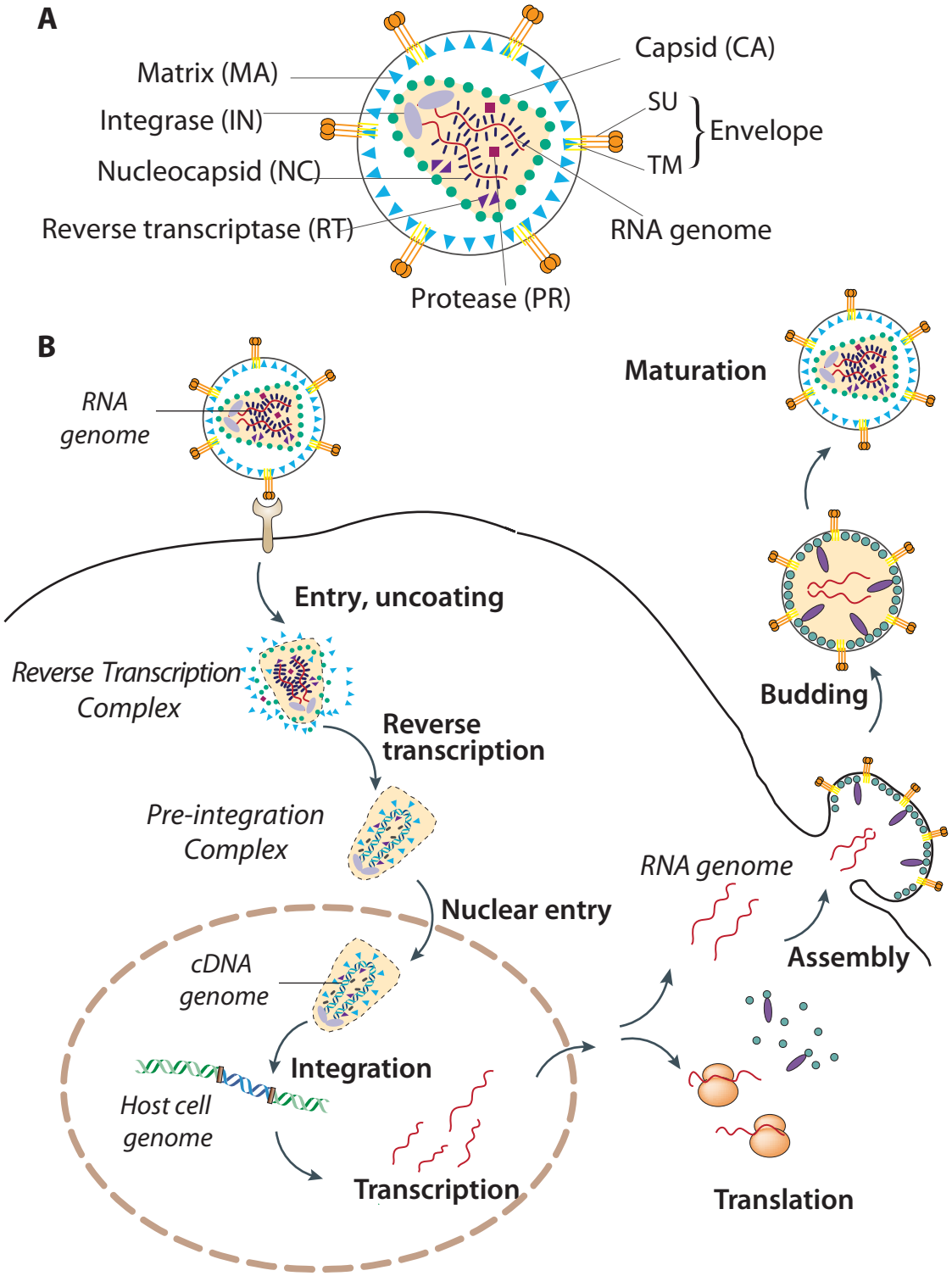

Figure 1-1. The retroviral lifecycle. A. A retroviral particle. **B.** The retroviral lifecycle. Adapted from Greene and Peterlin [9].

through the cytoplasm to the nucleus by interactions with actin microfilaments [11] and microtubules [12]. Further changes in core structure following reverse transcription generate the pre-integration complex (PIC), made up of the reverse transcribed genome, associated with a number of viral and cellular proteins [10, 13-15]. When the PIC reaches the host chromatin, the viral cDNA is integrated into the host genome by IN [16-20]. Integrated viral genomes, known as proviruses, then function akin to endogenous host genes, transcribed and translated by host machinery to generate new viral RNA genomes and virion proteins.

The genomes of retroviruses contain four essential genes encoding virion proteins: Gag, Pol, Pro and Env. Pro (PR) is the viral protease, and post-translationally cleaves polyproteins Gag and Pol during virion maturation. Gag is cleaved into the structural proteins MA, CA and NC, and in some cases additional proteins. Pol is cleaved into viral enzymes RT and IN (and in some viruses, PR). Env is cleaved into the surface glycoprotein (SU) and transmembrane protein (TM), which decorate the viral envelope (Env is cleaved by a cellular protease in the Golgi rather than the viral PR). Some retroviruses, known as complex retroviruses, additionally encode accessory proteins, which generally have roles in regulating viral gene expression, combating host defenses, or increasing infectivity [6]. All retroviral genomes contain regulatory regions at their 5' and 3' ends made up of U5 and U3 respectively, each flanked by a direct repeat, R. Following reverse transcription, these untranslated elements become rearranged and duplicated, so the cDNA and provirus contain a directly repeated sequence at each end, a long terminal repeat (LTR), in the

configuration U3-R-U5 [21]. The arrangement of the reverse-transcribed genome of HIV is shown in Figure 1-2A. The LTRs contain sequence elements required for IN function [22-24] and viral gene expression. U3 contains promoter and enhancer elements, including host transcription factor binding sites [25]. Sequences in R or U3 in the 3' LTR form the polyadenylation signal [26].

Virion proteins are translated as Gag, Gag-Pol, Env and Pro (Pro is expressed as part of Gag or Pol, or individually depending on the virus). Gag, Gag-Pol and Pro associate with the nascent viral genome to assemble the virion core. Env is targeted to the cell membrane, from where new virions bud. Budding envelops the core in cellular lipid membrane studded with the Env protein. Following budding, the virion polyproteins are processed by the viral protease, leading to structural changes to the core known as maturation.

The family *Retroviridae* is divided into two sub-families, the *orthoretrovirinae* and the *spumaretrovirinae*. The *orthoretrovirinae* have been better studied, since several members of the family cause disease in animal hosts. The *orthoretrovirinae* are classified into six genera on the basis of sequence similarity. The genera and example members of each genus are listed in Table 1-1. The lentiviruses, which will be the topic of this dissertation, are a family of complex retroviruses named for their 'slowness'. They infect a range of mammalian hosts, and infection is characterized by long incubation periods, persistent viral replication and the destruction of hematologic or immunologic cells [6]. This genus contains HIV, SIV and FIV, which cause immunodeficiency in human, simian and feline hosts

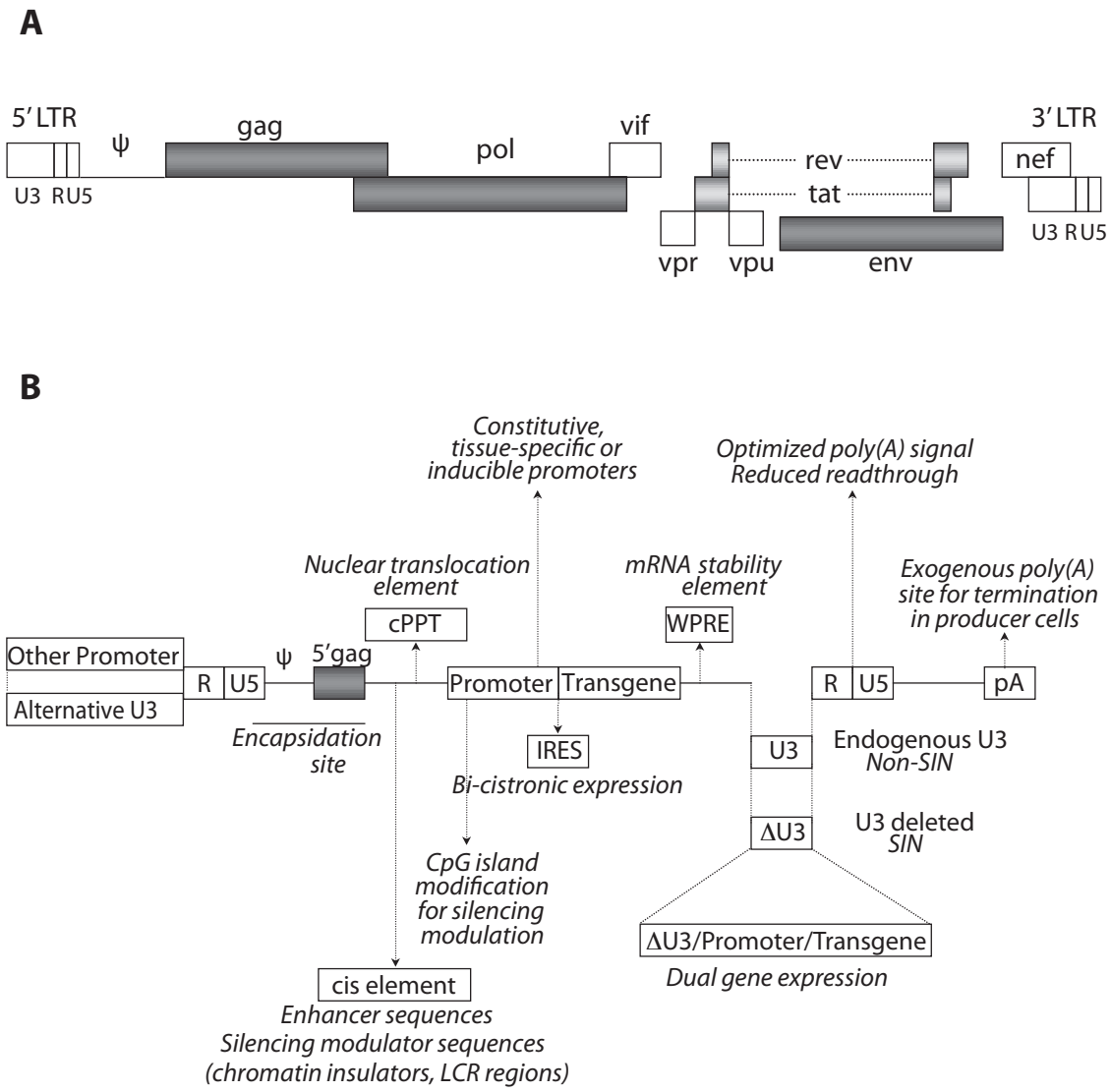

Figure 1-2. Lentiviral vectors. A. HIV genome map. **B.** Essential and optimized elements in a lentiviral gene transfer vector. cPPT, central polypurine tract; WPRE Woondchuck heptatits virus post-transcriptional responsive element; IRES, internal ribosome entry site. Adapted from Delenda [147].

Genus	Type species	Other examples	Notes
Alpharetrovirus	Avian leukosis virus (ALV)	Rous sarcoma virus (RSV)	Simple genome; associated with anemia, sarcoma and other tumors in avian hosts
Betaretrovirus	Mouse mammary tumor virus (MMTV)	Jaagsiekte sheep retrovirus (JSRV)	Simple genome; associated with cancer in mammalian hosts
Gammaretrovirus	Murine leukemia virus (MLV)	Feline leukemia virus (FeLV), Xenotropic murine leukemia virus-related virus (XMRV)	Simple genome; associated with cancer in mammalian hosts
Deltaretrovirus	Bovine leukemia virus (BLV)	Human T-lymphotropic virus (HTLV)	Complex genome; associated with leukemia and lymphoma in mammalian hosts
Epsilonretrovirus	Walleye dermal sarcoma virus (WDSV)		Complex genome; Infect fish
Lentivirus	Human immunodeficiency virus (HIV)	Simian immunodeficiency virus (SIV), Feline immunodeficiency virus (FIV), Equine infectious anemia virus (EIAV)	Complex genome; slow, chronic viral replication in mammalian hosts; can infect non-dividing cells

Table 1-1. Retroviral genera and example members. Adapted from Coffin *et al.* [6].

respectively, and EIAV, which causes anemia in horses. A distinguishing feature of lentiviruses is their ability to infect non-dividing cells, engendering great interest in steps of their lifecycle leading up to their interaction with the host genome [27], and enabling their use as gene delivery vectors in terminally differentiated cell types (discussed in section 1.9 of this chapter).

1.2 Clinical relevance of lentiviruses

The retrovirus of greatest clinical importance is arguably the complex lentivirus, human immunodeficiency virus (HIV), identified in 1983 as the causative agent of acquired immunodeficiency disorder syndrome (AIDS) [28, 29]. HIV is transmitted by direct sexual contact, contact with blood or blood products, and from mothers to infants intrapartum, peripartum or through breast-feeding [6]. Both cell-free and cell-associated virus can play a role in transmission [30]. HIV infects immune cells expressing the receptor CD4, namely T-cells and macrophages. Primary infection is associated with an acute phase of mononucleosis-like illness, high viremia and low CD4+ T-cell count 3-6 weeks after transmission [6]. Following acute infection, viral load drops and CD4+ count recovers, but over a period of several years, untreated infection results in a slow increase in viral load and decline in CD4+ T-cells. This period of slow decline is known as clinical latency, and is asymptomatic. Eventually, CD4+ T-cell numbers decline below a critical threshold, leading to AIDS: increased susceptibility to opportunistic infections, which ultimately lead to the patient's death. According to the World Health Organization,

an estimated 33.4 million people are currently infected with HIV worldwide and 2 million HIV-related deaths occurred in 2008 [31].

While an effective HIV vaccine remains elusive, pharmacological inhibitors of several steps of the HIV lifecycle have been developed and are used for the treatment of HIV. Roughly 30 inhibitors have been developed and approved for use in patients, targeting the entry, reverse transcription, integration or maturation steps of the viral lifecyle. These are typically administered as highly active antireroviral therapy (HAART), a cocktail of three drugs taken together to reduce the risk of development of resistance to any one drug. While the drugs currently in use have made a remarkable impact on HIV mortality and morbidity, they must be taken indefinitely and patients frequently develop resistance to them [32, 33], necessitating periodic changes to the patient's drug regimen. There therefore remains much interest in shedding light on the lifecycle of HIV and related lentiviruses, particularly interactions of the virus with host cell factors, to identify potential novel therapeutic targets.

1.3 Integration

This dissertation focuses on the integration step of the viral lifecycle. The ability of retroviruses to integrate into the host genome is responsible for several of the most challenging aspects of HIV treatment and eradication. Integration of the genome allows the establishment of latency, prevents complete elimination of viruses with antiretrovirals [34], and enables archiving of drug resistance mutations [35]. The covalent insertion of viral sequence into the host genome has also contributed to

genome evolution: roughly 8% of the human genome, for example, is composed of retroviral sequences and it is thought that some of these elements have been co-opted for the benefit of the host cell [36]. Finally, as described in detail in section 1.7 of this chapter, the ability of retroviruses to integrate DNA between their LTRs into host DNA has enabled their use in gene therapy, but the possibility of this event disrupting host gene expression is also the cause of safety concerns.

The integration reaction is illustrated in Figure 1-3. It is composed of three main steps: processing of viral DNA ends, joining of viral to host DNA, and repair of gaps. The viral enzyme integrase (IN) has been shown to be sufficient for the first two steps *in vitro* [19, 20, 37, 38]. IN removes two nucleotides from the 3' termini of the viral DNA, leaving recessed 3' hydroxyl groups [39, 40]. IN then catalyzes attack by these hydroxyl groups on phosphodiester bonds in the target DNA backbone [19, 20, 41]. This leaves single-stranded gaps between the points of joining on the two strands, which are probably repaired by host DNA repair enzymes [42].

The integrase protein is proteolytically cleaved from the Gag-Pol precursor upon virion maturation. IN is composed of three domains: the N-terminal zinc-binding domain (amino acids 1-50), the catalytic core (amino acids 50-212), and the C-terminal DNA-binding domain (amino acids 212-288) [43]. Three conserved amino acids ($D,DX_{35}E$, referred to as the catalytic triad) in the catalytic core domain (CCD) are required for catalysis [44-46]. The structure of the CCD is an RNase H-like fold, conserved among members of the RNase H phosphotransferase enzyme family [45, 47, 48]. The catalytic triad is brought together in space, and coordinates

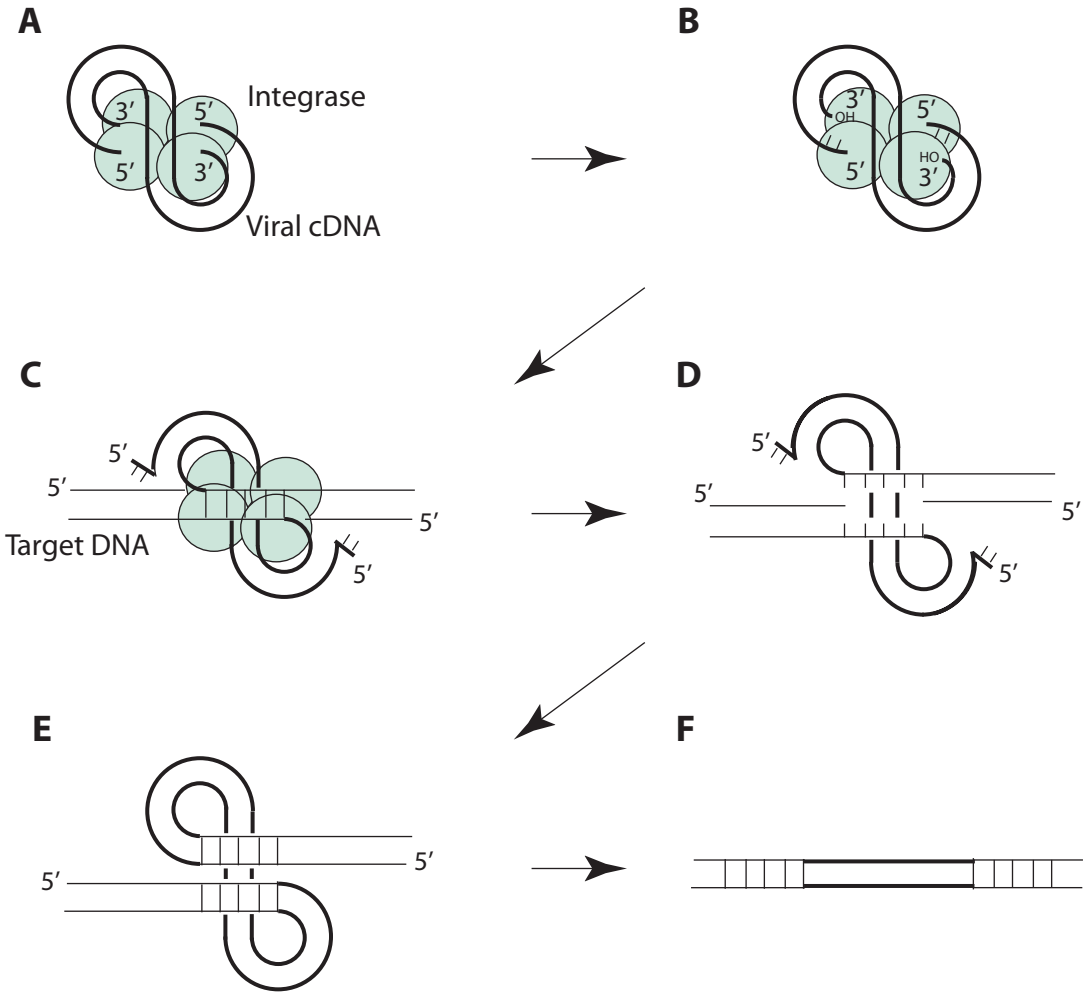

Figure 1-3. The integration reaction. **A.** Reverse-transcribed viral DNA (bold lines) is associated with integrase (IN), probably as a tetramer (green circles), in the pre-integration complex. **B.** Terminal cleavage. IN removes 2bp from the 3' ends of the viral DNA, exposing hydroxyl groups and leaving 5' overhangs. **C.** Strand transfer. IN catalyzes nucleophilic attack by the two 3' OH groups on the phosphate backbone at two positions in the host DNA (fine lines). The interval between the two positions differs by retrovirus (eg. 5bp for HIV). **D.** The reaction intermediate contains unpaired gaps at each viral-host DNA junction. **E.** Gap repair. Host enzymes are thought to fill in the gaps. **F.** The integrated provirus is identical in sequence to the reverse transcribed genome, and is flanked by a repeat in the host genome resulting from gap repair (in the case of HIV, 5bp). Adapted from Ciuffi and Bushman [189].

two divalent metal cations [49-53]. The core is also thought to function in DNA binding [24, 54] and contribute to target DNA sequence preference *in vitro* [55, 56]. The N- and C-terminal domains are thought to promote DNA binding and multimerization [57-63]. No crystal structure of full-length IN has been published yet, but two-domain structures show CCDs associated as dimers [64-67], leading to a model of IN function *in vivo* as a dimer of dimers [68].

In vivo, integration is carried out by the PIC, a nucleoprotein complex derived from the viral core [10, 13, 69]. PICs can be isolated from infected cells and carry out coordinate integration of both ends of endogenous cDNA into target DNA supplied *in vitro* [10, 70, 71]. HIV PICs have been shown to contain, as well as IN, the viral proteins MA, RT, NC and Vpr [15] but very little CA [14]. A number of host proteins have also been found to associate with purified IN or with the PIC, and the contribution of such host factors to integration is discussed below and is a focus of this dissertation.

1.4 Host factors in integration

Attempts to identify candidate integration cofactors have used a number of approaches. One approach has been to search for host proteins that stimulate integration by pre-integration complexes. It was observed that gel-filtration of PICs in the presence of high salt resulted in a loss of integrase activity, which could be restored by adding back cell extracts. By fractionating such extracts a number of host factors have been identified, including the non-histone chromatin protein HMGA1 in HIV PICs [72] and the chromatin-associated protein BAF in HIV [73] and MLV PICs

[74]. However, further studies of HMGA1 suggested it is not strictly required in HIV infection [75] and the role of BAF in infection remains unclear [76, 77]. Another approach has been identifying IN binding partners by yeast-two-hybrid. The SWI/SNF chromatin remodeling protein Ini1, was identified as interacting with HIV IN [78], though its role in infection now appears to be in assembly [79]. More recently, TNPO3/transportin-SR2 was identified by yeast-two-hybrid as interacting with HIV IN [80], and appears to be an essential nuclear import factor, though it may also function through CA binding [81]. Similarly, the transcriptional coactivator LEDGF/p75 was identified by its co-immunoprecipitation with IN overexpressed in human cells [82]. LEDGF/p75 appears to be an essential lentiviral integration tether and is discussed in detail below and in Chapters 2 and 3 of this dissertation. Most recently, genome-wide siRNA screens [83-85] have identified hundreds of candidate host factors necessary for HIV infection. Potential roles of some of these factors in integration targeting are discussed in Chapter 5 of this dissertation.

1.5 Genomic distribution of integration sites

Early studies of integration *in vitro* suggested any DNA sequence could serve as a target for integration by purified IN or PICs. Relatively weak local sequence preference [86-89], and a preference for distorted nucleosome-associated DNA [86, 90-92] were observed. The advent of genome sequencing enabled genome-wide studies of integration in cells, which have shown clear biases in the distribution of retroviral integration sites with respect to various genomic features [93-95]. Indeed different genera of retroviruses show different integration site preferences. In a

number of cell types, HIV [93, 96-98] and other lentiviruses [99, 100] favor transcription units as integration sites, particularly active genes. In contrast, gammaretroviruses such as MLV and XMRV show a preference for CpG islands and gene 5' ends [95, 101], while alpharetroviruses such as ASLV, betaretroviruses such as MMTV and deltaretroviruses such as HTLV show relatively random integration patterns, with weak or no favoring of transcription units [94, 96, 102, 103].

The weak sequence preferences demonstrated by retroviral integrases do not fully account for the genomic distribution of integration sites [104]. Hypotheses that have been proposed to explain integration site targeting by retroviruses center around the ideas of chromatin accessibility or tethering. Since much of the DNA in mammalian cells is tightly wrapped into higher order chromatin structures, and these structures change with transcriptional status and cell cycle phase, it may be that integration can only occur in regions that are in an exposed conformation. The bias of MLV integration toward DNase I hypersensitive sites [105, 106] and of HIV away from alphoid repeats located in pericentric heterochromatin [89] lend some support to this idea. However, the distinct patterns of different retroviruses in the same cell types are suggestive of virus-specific tethers, rather than simple accessibility.

Consistent with the tracking of integration site preferences with retroviral genus, viral elements have been shown to determine integration site distributions. In a study of chimeric HIV viruses bearing MLV IN, Gag or both, viruses containing MLV IN were found to integrate with an MLV-like distribution, with further influence of Gag when both MLV proteins were present [107]. The idea currently

favored in the field is therefore that retroviral PICs are targeted to particular regions of the host chromatin through interactions between viral and host proteins. Tethering interactions are well documented in yeast retrotransposons, which are closely related to retroviruses [108, 109], and artificial fusions of HIV integrase to sequence-specific DNA-binding domains have been shown to direct integration to their recognition sites [110, 111]. IN, and other viral proteins identified as determining integration site preference [107] are obvious candidate binding partners for cellular tethering factors, though any viral component of the PIC described above could potentially play a role.

1.6 LEDGF/p75 in lentiviral integration

LEDGF/p75 is a ubiquitously expressed nuclear protein now widely accepted as a cofactor for lentiviral integration. Interest in LEDGF/p75 in the HIV field began when it was found to interact with overexpressed HIV IN in the nucleus [82] and was identified as an IN binder by yeast-two-hybrid [112]. The protein was identified as p75, one of two splice variants from the PSIP1/LEDGF gene, reported to be a transcriptional coactivator that co-purified with the general transcription factor PC4 [113]. The cellular function of LEDGF/p75 remains unclear. In its initial isolation with PC4, it was shown to have weak transcriptional coactivator activity, though the p52 splice variant was more active. Around the same time, the protein was also isolated from lens epithelium cells [114], and given the name Lens Epithelium Derived Growth Factor. It has been implicated in cellular stress responses [115, 116] including apoptosis [117] and tumor angiogenesis [118], and it is disrupted in

chromosomal translocations associated with acute and chronic myeloid leukemias [119, 120].

Mice lacking LEDGF/p75 expression have been generated, either by gene trap disruption [121] or knockout [122]. The mice are viable, showing some perinatal death due to problems feeding, and a range of phenotypic abnormalities in adulthood, including low fertility and homeotic defects, but normal lens epithelia. Murine embryonic fibroblasts (MEFs) derived from these mice, as well as human cell lines stably knocked down for LEDGF/p75, grow normally and were used in the experiments described in later chapters.

Figure 1-4 illustrates the domain structure of LEDGF/p75. At its N-terminus (amino acids 1-325) is a PWWP domain, a nuclear-localization signal, two AT-hook motifs and three charged regions, which have all been implicated in chromatin binding [123-125]. The PWWP domain (named for its Pro-Trp-Trp-Pro motif) is found in a number of chromatin-binding proteins, and is thought to be a member of the Tudor domain Royal family [126-128]. LEDGF/p75 shows sequence-non-specific DNA binding [124], but the relative contributions of binding to DNA and protein in chromatin remain unclear.

p75's C-terminus contains a domain that binds lentiviral integrases, but not the IN of other classes of retroviruses [123, 129-131], named the integrase-binding domain (IBD, amino acids 347-429). On IN, the CCD is minimally sufficient for the interaction, but is bolstered by contacts with the NTD [112, 132]. Crystal structures have been solved of the LEDGF/p75 IBD bounds to the HIV IN CCD and NTD, and

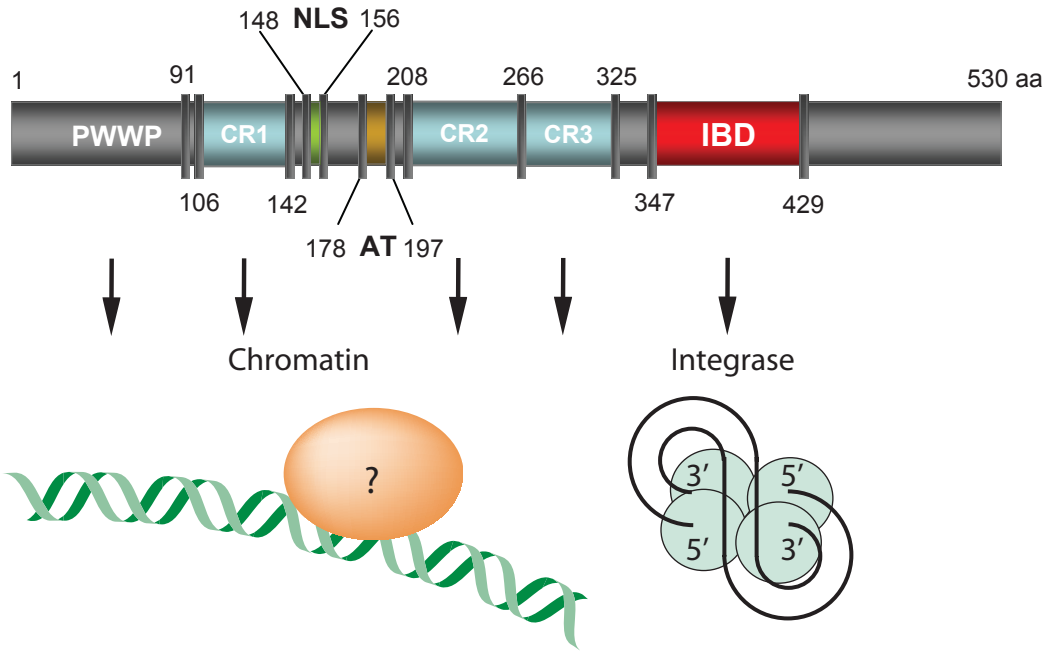

Figure 1-4. LEDGF/p75 domain structure and interactions. PWWP, PWWP domain; CR, charged region; NLS nuclear localization signal; AT, AT-hooks; IBD, integrase-binding domain. Arrows represent interaction of various domains with chromatin (? represents unknown chromatin ligands, the helix represents DNA) or integrase, represented as a tetramer bound to viral cDNA. Domain structure adapted from Gijsbers *et al.* [190].

show each LEDGF/p75 IBD interacting with two CCD interfaces and one NTD [68, 133]. A number of cellular proteins have been also been shown to interact with LEDGF/p75's IBD: JPO2, a c-myc-binding protein [134]; pogZ, a domesticated transposase with sequence similarity to retroviral IN [135]; and menin, an adaptor protein associated with the histone-methyltransferase MLL [136].

Consistent with its modular structure – the N-terminus binding chromatin and the C-terminus binding integrase – LEDGF/p75 has been shown to function as a molecular tether, recruiting IN (and its other binding partners) to chromatin. When IN was overexpressed in wild-type cells it colocalized with LEDGF/p75 and chromatin [82, 129, 132]. Depletion of LEDGF/p75 by RNAi or mutations that abrogated LEDGF/p75's interaction with IN or chromatin led to loss of IN nuclear localization [112, 129, 132, 137, 138]. This led to early proposals that LEDGF/p75 was responsible for nuclear import of IN [132]. However, subsequent studies with NLS-deleted LEDGF/p75 showed that IN could achieve nuclear localization in dividing cells due to nuclear-cytosolic mixing during division [123]. Whether LEDGF/p75 effects transport of IN across the nuclear membrane or acts to retain it on chromatin after import is not fully determined, though experiments with whole virus favor the latter model (see below).

LEDGF/p75 enhances IN activity *in vitro* [82]. This stimulatory activity requires integrase binding and DNA binding [124, 131]. This may be a biologically important activity of LEDGF/p75, but caution is warranted in this interpretation because many nonspecific DNA binding proteins display this activity *in vitro*.

LEDGF/p75 also enhances the solubility of IN [130], and protects it from proteasomal degradation, independently of its chromatin-tethering function [139, 140].

In addition to the above evidence of LEDGF/p75's role in binding, trafficking and stimulation of purified or overexpressed IN, its role in cellular infection has also been demonstrated. It has been suggested to be a component of functional PICs [129] and partly reconstitute the activity of salt-disrupted PICs [141]. Early studies of LEDGF/p75 knockdown reported little [142, 143] or no effect [129, 141, 144] on the level of HIV infectivity. However, a role as an integration cofactor was supported by the finding that a virus bearing an IN mutation that disrupted LEDGF/p75 binding while preserving catalytic activity exhibited a severe infection block [112], though effects on other infection steps were hard to rule out. Similarly, overexpression of the IBD was reported to act as a dominant negative and inhibit infection at the step of integration [138], arguing that LEDGF/p75 had a role in HIV integration *in vivo*.

Based on this model, it seemed probable that LEDGF/p75 could be a determinant of the genomic distribution of lentiviral integration events. Indeed, a role for LEDGF/p75 in integration site selection was demonstrated in human cell lines stably expressing siRNAs against LEDGF/p75 [144]. Knockdown led to a significant, but partial, reduction in the frequency of integration in transcription units, specifically in LEDGF-responsive genes (as determined by transcriptional profiling of knockdown cells) and A/T-rich regions of the genome, hypothesized to be bound by LEDGF by virtue of its AT-hook motifs.

The absence of an infectivity defect in knockdown cells generated by a variety of labs was ultimately argued to be due to residual chromatin-associated LEDGF/p75 expression. Intensified knockdown cells were engineered by stable expression of lentiviral vectors containing LEDGF/p75 shRNA constructs [140]. In these cells, no residuum was detected and HIV infection was reduced 31-fold at the integration step of the viral lifecycle. An increased accumulation of 2-LTR circles was noted, a species of viral cDNA known to be generated by nuclear NHEJ enzymes and thus taken as a measure of nuclear import [145, 146]. Rescue of the infection defect required both chromatin- and integrase-binding capabilities of LEDGF/p75. The LEDGF/p75 status of the producer cell had no impact on HIV infectivity, arguing that LEDGF/p75 is not packaged into virions.

At the time this dissertation research was started, integration site selection in intensified knockdown cells had not been studied. Additionally, a gene-trap LEDGF/p75 mouse model had been generated, in which lentiviral infection had not been studied. This left open the possibility that the remaining targeting to transcription units observed by Ciuffi and colleagues [144] was due to the residual LEDGF/p75 expression in the cells used, and additionally provided us another model in which to verify the effect of LEDGF/p75 depletion on lentiviral infection. It remained unclear:

1. whether LEDGF/p75 was necessary for lentiviral integration site selection in full
2. whether it cooperated with or was antagonized by other determinants
3. whether it was sufficient to determine the sites of lentiviral integration.

Addressing these questions was therefore one goal of the work described in this dissertation. In Chapter 2, we seek to determine the necessity of LEDGF/p75 by studying HIV and EIAV integration sites in intensified human knockdown cells and MEFs from a gene-trap-disrupted mouse model with undetectable LEDGF/p75 expression. In Chapter 3, we address the issue of sufficiency by using fusion proteins containing the LEDGF/p75 IBD and alternative chromatin binding domains to retarget lentiviral integration. In Chapters 5 and 6, we consider other factors that may contribute to integration site selection.

Another part of this dissertation concerns integration in gene therapy, and we now turn to that topic.

1.7 Use of retroviral vectors for gene therapy

The ability of retroviruses to covalently integrate their genetic material into the host genome has made them attractive vehicles for the delivery of corrective genes in patients suffering from genetic diseases. Generation of a retroviral gene delivery vector essentially involves replacing the viral genes *gag*, *pol* and *env* with the host transgene. This is illustrated in Figure 1-2B. The resulting transfer vector, bearing the transgene, viral packaging signal and LTR sequences, functions like a retroviral genome. The packaging proteins (Gag, Gag-Pol and Env) must be provided in *trans*, since they are absent from the 'genome' that bears the transgene. Cells therefore express, transiently or stably, the transfer vector, a Gag-Pol vector and an envelope vector, and produce packaged virions [147, 148]. These virions are used to transduce patient cells, the transfer vector undergoes reverse transcription and

integration, and the corrective gene is stably expressed. Retroviruses have a coding capacity of around 7-7.5kb, can be manufactured at relatively high titers, and have a fairly high transduction efficiency, making them practical transfer vectors [149-151].

Gammaretroviral vectors (based on MLV) were the first to be developed, and remain the most widely used, with 333 clinical trials completed or in progress worldwide as of December 2009 [152]. A major limitation of gammaretroviral vectors is their inability to infect non-dividing cells. In an effort to expand the range of disorders to which gene therapy could be applied, lentiviral vectors, which are able to infect non-dividing cells, have recently increased in popularity, with 24 trials currently completed or on-going.

The greatest success has been in the development of techniques for treatment of hematological disorders such as SCID-X1, ADA-SCID and CGD, ALD and β-thalassemia. In these protocols, bone marrow is harvested from the patient and transduced *ex vixo*. Transduced cells are then transplanted back into the patient following myeloablation, and gene-corrected stem cells reconstitute the bone marrow [151].

The work of many labs has enhanced the design of transfer vectors to improve transgene expression and prevent silencing. Some of these elements are illustrated in Figure 1-2B. Transduction efficiency is improved by the incorporation of sequence elements to enhance RT and possibly PIC nuclear import, for example a polypurine tract in the cDNA, the cPPT [153]. Replacement of the viral LTR promoter with an alternative cellular promoter can enhance transcription initiation, target certain cell

types and enhance long-term expression [147]. Transcript nuclear export can be maximized by incorporating an intron or a post-transcriptional regulatory element such as the WPRE from Woodchuck Hepatitis virus [154, 155]. The risk of transgene silencing by DNA methylation is reduced by incorporating insulator elements that prevent the spread of epigenetic modifications from surrounding DNA [147, 148]. A range of envelope glycoproteins are now also in use, enabling some degree of tissue-specific targeting [156].

One safety concern over the use of retroviral vectors has been the potential for reconstitution of an infectious retrovirus by recombination of transfer and packaging vectors. This has been dealt with by separating *gag-pol* and *env* sequences onto two separate packaging plasmids that do not contain overlapping viral sequence elements, reducing the probability of recombination during vector production [157]. Another safety issue is the possibility of insertional activation. This is discussed in detail below, and remains a serious concern, despite various modifications to vector design intended to mitigate it.

1.8 Insertional activation

Since their discovery, retroviruses have been implicated in carcinogenesis [158, 159]. Indeed studies of tumor-associated retroviruses have contributed to our understanding of the development of cancer [6]. Retroviruses can exert oncogenic effects by encoding an oncogene, either a captured cellular gene [159-161] or a modified viral factor with oncogenic properties [162]. Alternatively, retroviruses lacking an encoded oncogene can alter the expression of a host growth control gene

close to the site of proviral integration by insertional activation.

Insertional activation can be caused by a number of mechanisms [6]. One is upregulation of transcription of an oncogene by retroviral promoter or enhancer sequences inserted a short distance upstream of the gene. Alternatively, proviruses may integrate within a gene, resulting in transcriptional readthrough, forming a hybrid transcript of viral and host sequence. This hybrid may act as an aberrantly active growth factor, for example encode a constitutively active oncogene missing a regulatory domain. Finally, an integrated provirus may separate a growth-control gene from non-coding regions that modulate its expression. Tumors arising by insertional activation usually have a long latency – assuming the altered locus has a dominant phenotype, the initial integration event may impart a growth advantage on the cell, but additional mutations (second and third 'hits') will likely be required for a tumor to develop. Rarely, retroviruses can promote transformation by inactivating a tumor suppressor gene, though in this case the other allele must be inactivated as well.

In a number of clinical trials of retroviral vectors, insertional activation has resulted in adverse events. In the SCID-X1 trial, for example, 5 of 19 children treated with a gammaretroviral vector containing the common cytokine receptor γc chain went on to develop leukemia [163-165]. Analysis of the genomic sites of vector integration can provide evidence of insertional activation and shed light on the mechanism of oncogenesis. Integration events that alter the expression of cellular growth control genes would be expected to impart a growth advantage to the cells

harboring them. Those cells would therefore accumulate in the treated individual, and be more frequently recovered upon random sampling of circulating cells and bone marrow. Such analysis has been carried out for a number of gene therapy trials. In the SCID-X1 trial, for example, samples from patients who developed leukemia exhibited integration sites within or near the known growth-control genes LMO2, BMI1 and CCND2 in blast cells [164]. Clonal dominance was also observed in a trial of a gammaretroviral vector administered for the treatment of chronic granulomatous disease, which progressed to leukemia. In this case, both patients developed a clonal expansion of myeloid cells bearing integration sites in MDS1/EVI1, PDRM16 or SETBP1 and myelodysplasia [166, 167].

The factors determining the incidence and consequences of insertional activation are not fully understood, but are likely a combination of vector regulatory elements, the nature of the transgene, the culture and transduction conditions employed and characteristics of the target cell [168].

The contribution of cell-intrinsic factors to the incidence of insertional activation is relatively poorly characterized. The self-renewal properties of the target cell likely affect the consequences of vector integration. For example, Recchia and colleagues reported [169] that gammaretroviral transduction of terminally differentiated T-cells, though altering the expression of a large number of cellular genes, did not result in clonal skewing as long as 9 years post-transplantation. Similarly, Kustikova *et al.* reported that clonal dominance developed following gammaretroviral transduction of hematopoietic stem cell populations, but not more

lineage-restricted progenitors [170].

Likewise, the nature of the transgene and the nature of the disorder being treated is thought to affect the potential for insertional activation. For example, in the SCID-X1 trial, it is likely that the fact the γc chain was required for the survival of the targeted cells and that corrected cells expanded to fill an empty hematological compartment meant that transduced cells already had a growth advantage, increasing the selective forces driving clonal outgrowth [168].

The determinant of insertional activation viewed as the most straightforward to control is vector design. One proposed approach to reducing the risk of gene therapy would be to target integration events to specific sites in the genome, chosen to lie far from growth-control genes to minimize the risk of insertional activation. This is not yet a practical approach, but some success has been achieved creating chimeric proteins to retarget integration. Chapter 3 of this dissertation describes retargeting of lentiviral integration out of transcription units using such a LEDGF/p75 fusion.

In the absence of targeted integration, numerous vector design modifications have been proposed to reduce the vector's impact on the expression of nearby genes. Notably, many of these features were not present in the vectors used in the clinical trials and resulting adverse events described above [166, 171]. Use of physiologic cellular promoters such as PGK or EF1α to drive transgene expression, rather than strong retroviral promoters, has been shown to reduce transactivation [172]. The inclusion of insulator elements in vectors, in addition to reducing transgene silencing,

reduces activation of neighboring genes by proviral promoter and enhancer elements [173, 174]. Deletion of U3, which contains the viral promoter sequence, from the 3' LTR of the viral genome reduces transactivation of neighboring genes [175, 176]. Studies of insertional activation with such vectors, termed 'self-inactivating' (SIN), have supported the idea that they are less genotoxic [172, 177], though there remain examples of tumor development with SIN vectors [173, 178]. Insertional activation by leaky vector transcription can be reduced by the incorporation of exogenous polyadenylation signals such as that of SV40 in addition to that in the 3' U5 [179]. Finally, lentiviral vectors have been proposed to be safer than gammaretroviral vectors, as discussed below.

1.9 Lentiviral vectors

Until 2008, only gammaretroviral vectors had been used in clinical trials of gene therapy, though lentiviral vectors have long been attractive due to their ability to transduce non-dividing cells. Additionally, it was expected that lentiviral vectors might have a better safety profile. Unlike gammaretroviruses, insertional oncogenesis is not a common feature of infection with lentiviruses, for example HIV. Though it has been reported [180], none of the data to date is convincing. It has also been posited that differences in integration site preferences between letiviruses and gammaretroviruses might impact their safety profiles [148, 181]. Gammaretroviral vectors show a strong propensity to integrate at promoters and gene 5' ends [95], where transcriptional read-through from the viral LTR can lead to upregulation of the downstream gene [182]. Lentiviruses, on the other hand, favor integration in the

bodies of transcription units, avoiding regulatory 5' regions [93, 94, 98]. Indeed, studies in tumor-prone mouse models [183, 184] and tissue culture systems [185] have reported less genotoxicity resulting from lentiviral than gammaretroviral transduction.

Three clinical trials have been conducted using lentiviral vectors in humans, and have yielded mixed evidence regarding the consequences of integration. The first trial involved delivery of an HIV *env* antisense payload to terminally differentiated T-cells infected with HIV [186], and integration events in these patients showed no evidence of enrichment of sites in proto-oncogenes following transduction [187]. The second published trial, to treat ALD, involved delivery with a SIN vector containing the ABCD1 gene into hematopoietic stem cells [188]. Integration sites in these two patients also showed sustained polyclonality up to 24 months after transplantation, and clear clinical benefit was achieved. Thirdly, one patient was treated for β-thalassemia with a SIN lentiviral vector encoding β-globin (Cavazzana-Calvo *et al.*, submitted), and again clinical benefit was achieved, though in this case a clonal expansion bearing a site within the proto-oncogene HMGA2 was observed.

Given our limited understanding of the factors determining the consequences of retroviral gene therapy, the results of the human β-thalassemia trial raised questions about the possibility of SIN lentiviral vector integration near growth-control genes imparting a selective advantage and leading to preferential outgrowth of the target cell. In Chapter 4, we present a study of the distribution of integration sites from the same lentiviral vector as was used in the human β-thalassemia trial, used to

treat a mouse model. We sought to determine the generality of the finding in the human trial and contribute to our understanding of the consequences of integration with this vector for the infected cell.

1.10 References

1. Brindley MA, Maury W. (2005) Endocytosis and a low-pH step are required for productive entry of equine infectious anemia virus. J Virol 79(23): 14482-14488.

2. McClure MO, Sommerfelt MA, Marsh M, Weiss RA. (1990) The pH independence of mammalian retrovirus infection. J Gen Virol 71 (Pt 4)(Pt 4): 767-773.

3. Mothes W, Boerger AL, Narayan S, Cunningham JM, Young JA. (2000) Retroviral entry mediated by receptor priming and low pH triggering of an envelope glycoprotein. Cell 103(4): 679-689.

4. Stein BS, Gowda SD, Lifson JD, Penhallow RC, Bensch KG, et al. (1987) pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. Cell 49(5): 659-668.

5. Gomez C, Hope TJ. (2005) The ins and outs of HIV replication. Cell Microbiol 7(5): 621-626.

6. Coffin JM, Hughes SH, Varmus HE. (1997) Retroviruses. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

7. Goila-Gaur R, Strebel K. (2008) HIV-1 vif, APOBEC, and intrinsic immunity. Retrovirology 5: 51.

8. Kleiman L, Halwani R, Javanbakht H. (2004) The selective packaging and annealing of primer tRNALys3 in HIV-1. Curr HIV Res 2(2): 163-175.

9. Greene WC, Peterlin BM. (2002) Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. Nat Med 8(7): 673-680.

10. Brown PO, Bowerman B, Varmus HE, Bishop JM. (1987) Correct integration of retroviral DNA in vitro. Cell 49: 347-356.

11. Bukrinskaya A, Brichacek B, Mann A, Stevenson M. (1998) Establishment of a functional human immunodeficiency virus type 1 (HIV-1) reverse transcription complex involves the cytoskeleton. J Exp Med 188(11): 2113-2125.

12. McDonald D, Vodicka MA, Lucero G, Svitkina TM, Borisy GG, et al. (2002) Visualization of the intracellular behavior of HIV in living cells. J Cell Biol 159(3): 441-52.

13. Bowerman B, Brown PO, Bishop JM, Varmus HE. (1989) A nucleoprotein complex mediates the integration of retroviral DNA. Genes and Development 3: 469-478.

14. Farnet CM, Haseltine WA. (1991) Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. J. Virol. 65: 1910-1915.

15. Miller MD, Farnet CM, Bushman FD. (1997) Human immunodeficiency virus type 1 preintegration complexes: Studies of organization and composition. J. Virol. 71: 5382-5390.

16. Donehower LA, Varmus HE. (1984) A mutant murine leukemia virus with a single missense codon in pol is defective in a function affecting integration. Proc. Natl. Acad. Sci. USA 81: 6461-6465.

17. Schwartzberg P, Colecilli J, Goff SP. (1984) Construction and analysis of deletion mutations in the pol gene of moloney murine leukemia virus: A new viral function required for productive infection. Cell 37: 1043-1052.

18. Panganiban AT, Temin HM. (1984) The retrovirus pol gene encodes a product required for DNA integration: Identification of a retrovirus int locus. Proc. Natl. Acad. Sci. USA 81: 7885-7889.

19. Craigie R, Fujiwara T, Bushman F. (1990) The IN protein of moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell 62: 829-837.

20. Bushman FD, Fujiwara T, Craigie R. (1990) Retroviral DNA integration directed by HIV integration protein *in vitro*. Science 249: 1555-1558.

21. Gilboa E, Sudha WM, Goff S, Baltimore D. (1979) A detailed model of reverse transcription and tests of crucial aspects. Cell 18: 93-100.

22. Bushman FD, Craigie R. (1990) Sequence requirements for integration of moloney murine leukemia virus DNA *in vitro*. Journal of Virology 64: 5645-5648.

23. Leavitt AD, Rose RB, Varmus HE. (1992) Both substrate and target

oligonucleotide sequences affect in vitro integration mediated by human immunodeficiency virus type 1 integrase protein produced in saccharomyces cerevisiae. J. Virol. 66: 2359-2368.

24. Esposito D, Craigie R. (1998) Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. EMBO J. 17: 5832-5843.

25. Pereira LA, Bentley K, Peeters A, Churchill MJ, Deacon NJ. (2000) A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. Nucleic Acids Res 28(3): 663-668.

26. Guntaka RV. (1993) Transcription termination and polyadenylation in retroviruses. Microbiol Rev 57(3): 511-521.

27. Suzuki Y, Craigie R. (2007) The road to chromatin - nuclear entry of retroviruses. Nat Rev Microbiol 5(3): 187-196.

28. Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, et al. (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science 220(4599): 865-867.

29. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, et al. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220(4599): 868-871.

30. Van de Perre P. (1999) Transmission of human immunodeficiency virus type 1 through breast-feeding: How can it be prevented? J Infect Dis 179 Suppl 3: S405-7.

31. UNAIDS. (2009) AIDS epidemic update 2009. .

32. Richman DD, Morton SC, Wrin T, Hellmann N, Berry S, et al. (2004) The prevalence of antiretroviral drug resistance in the united states. AIDS 18(10): 1393-1401.

33. Little SJ, Holte S, Routy JP, Daar ES, Markowitz M, et al. (2002) Antiretroviral-drug resistance among patients recently infected with HIV. N Engl J Med 347(6): 385-394.

34. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, et al. (1997) Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science 278: 1295-1300.

35. Wind-Rotolo M, Durand C, Cranmer L, Reid A, Martinson N, et al. (2009) Identification of nevirapine-resistant HIV-1 in the latent reservoir after single-dose nevirapine to prevent mother-to-child transmission of HIV-1. J Infect Dis .

36. Bushman FD. (2001) Lateral DNA transfer: Mechanisms and consequences. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

37. Katzman M, Katz RA, Skalka AM, Leis J. (1989) The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. J. Virol. 63: 5319-5327.

38. Sherman PA, Fyfe JA. (1990) Human immunodeficiency virus integration protein expressed in escherichia coli possesses selective DNA cleaving activity. Proc. Natl. Acad. Sci. USA 87: 5119-5123.

39. Fujiwara T, Mizuuchi K. (1988) Retroviral DNA integration: Structure of an integration intermediate. Cell 54: 497-504.

40. Brown PO, Bowerman B, Varmus HE, Bishop JM. (1989) Retroviral integration: Structure of the initial covalent complex and its precursor, and a role for the viral IN protein. Proc. Natl. Acad. Sci. U S A 86: 2525-2529.

41. Katz RA, Merkel G, Kulkosky J, Leis J, Skalka AM. (1990) The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. Cell 63: 87-95.

42. Yoder K, Bushman FD. (2000) Repair of gaps in retroviral DNA integration intermediates. J. Virol. 74: 11191-11200.

43. Engelman A, Bushman FD, Craigie R. (1993) Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. EMBO J. 12: 3269-3275.

44. Engelman A, Craigie R. (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function *in vitro*. J. Virol. 66: 6361-6369.

45. Kulkosky J, Jones KS, Katz RA, Mack JPG, Skalka AM. (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among Retroviral/Retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. 12: 2331-2338.

46. van Gent DC, Oude Groneneger AAM, Plasterk RHA. (1993) Identification of amino acids in HIV-2 integrase involved in site-specific hydrolysis and

alcoholysis of viral DNA termini. Nuc. Acids Res. 21: 3373-3377.

47. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, et al. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: Similarity to other polynucleotidyl transferases. Science 266: 1981-1986.

48. Nowotny M. (2009) Retroviral integrase superfamily: The structural perspective. EMBO Rep 10(2): 144-151.

49. Bujacz G, Jaskolski M, Alexandratos J, Wlodawer A, Merkel G, et al. (1996) The catalytic domain of avian sarcoma virus integrase: Conformation of the active-site residues in the presence of divalent cations. Structure 4(1): 89-96.

50. Gao K, Wong S, Bushman F. (2004) Metal binding by the D,DX35E motif of human immunodeficiency virus type 1 integrase: Selective rescue of cys substitutions by Mn2+ *in vitro*. J. Virol. 78(13): 6715-22.

51. Diamond TL, Bushman FD. (2006) Role of metal ions in catalysis by HIV integrase analyzed using a quantitative PCR disintegration assay. Nucleic Acids Res 34(21): 6116-25.

52. Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, et al. (1998) Three new structures of the core domain of HIV-1 integrase: An active site that binds magnesium. Proc. Natl. Acad. Sci. USA 95: 9150-9154.

53. Maignan S, Guilloteau JP, Zhou-Liu Q, Clement-Mella C, V M. (1998) Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: High level of similarity of the active site with other viral integrases. J. Mol. Biol. 282: 359-368.

54. Jenkins TM, Esposito D, Engelman A, Craigie R. (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. EMBO J. 16: 6849-6859.

55. Shibagaki Y, Chow SA. (1997) Central core domain of retroviral integrase is responsible for target site selection. J. Biol. Chem. 272(13): 8361-9.

56. Appa RS, Shin CG, Lee P, Chow SA. (2001) Role of the nonspecific DNA-binding region and alpha helicies within the core domain of retroviral integrases in selecting target DNA sites for integration. J. Biol. Chem. 276: 45848-45855.

57. Hazuda DJ, Wolfe AL, Hastings JC, Robbins HL, Graham PL, et al. (1994) Viral long terminal repeat substrate binding characteristics of the human immunodeficiency virus type 1 integrase. The Journal of Biological Chjemistry

269(6): 3999-4004.

58. Vincent KA, Ellison V, Chow SA, Brown PO. (1993) Characterization of human immunodeficiency virus type 1 integrase expressed in escherichia coli and analysis of variants with amino-terminal mutations. Journal of Virology 67(1): 425-437.

59. Zheng R, Jenkins TM, Craigie R. (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. Proc. Natl. Acad. Sci. USA 93: 13659-13664.

60. Woerner AM, Marcus-Sekura CJ. (1993) Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. Nucleic Acids Res. 21: 3507-3511.

61. Engelman A, Hickman AB, Craigie R. (1994) The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. J. Virol. 68: 5911-5917.

62. Andrake MD, Skalka A-. (1995) Multimerization determinants reside in both the catalytic core and C terminus of avain sarcoma virus integrase. J. Biol. Chem. 270: 29299-29306.

63. Jenkins TM, Engelman A, Ghirlando R, Craigie R. (1996) A soluble active mutant of HIV-1 integrase: Involvement of both the core and carboxyl-terminal domains in multimerization. J. Biol. Chem. 271: 7712-7718.

64. Yang ZN, Mueser TC, Bushman FD, Hyde CC. (2000) Crystal structure of an active two-domain derivative of rous sarcoma virus integrase. J. Mol. Biol. 296: 535-548.

65. Chen JC-, Krucinski J, Miercke LJW, Finer-Moore JS, Tang AH, et al. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding. Proc. Natl. Acad. Sci. USA 97: 8233-8238.

66. Chen Z, Yan Y, Munshi S, Li Y, Zugay-Murphy J, et al. (2000) X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50-293)--an initial glance of the viral DNA binding platform. J. Mol. Biol. 296: 521-533.

67. Wang JY, Ling H, Yang W, Craigie R. (2001) Structure of a two-domain fragment of HIV-1 integrase: Implications for domain organization in the intact protein. EMBO J 20: 7333-7343.

68. Hare S, Di Nunzio F, Labeja A, Wang J, Engelman A, et al. (2009) Structural basis for functional tetramerization of lentiviral integrase. PLoS Pathog 5(7): e1000515.

69. Ellison VH, Abrams H, Roe T, Lifson J, Brown PO. (1990) Human immunodeficiency virus integration in a cell-free system. J. Virol. 64: 2711-2715.

70. Farnet CM, Haseltine WA. (1990) Integration of human immunodeficiency virus type 1 DNA in vitro. Proc Natl Acad Sci U S A 87(11): 4164-4168.

71. Hansen MST, Smith GJI, Kafri T, Molteni V, Siegel JS, et al. (1999) Integration complexes derived from HIV vectors for rapid assays in vitro. Nat. Biotech. 17: 578-582.

72. Farnet C, Bushman FD. (1997) HIV-1 cDNA integration: Requirement of HMG I(Y) protein for function of preintegration complexes in vitro. Cell 88: 1-20.

73. Chen H, Engelman A. (1998) The barrier-to-autointegration protein is a host factor for HIV type 1 integration. Proc. Natl. Acad. Sci. USA 95: 15270-15274.

74. Lee MS, Craigie R. (1994) Protection of retroviral DNA from autointegration: Involvement of a cellular factor. Proc. Natl. Acad. Sci. USA 91: 9823-9827.

75. Beitzel B, Bushman FD. (2003) Construction and analysis of cells lacking the HMGA gene family. Nuc. Acids Res. 31: 5025-5032.

76. Jacque JM, Stevenson M. (2006) The inner-nuclear-envelope protein emerin regulates HIV-1 infectivity. Nature 441(7093): 641-645.

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

78. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP. (1994) Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science 266(5193): 2002-6.

79. Yung E, Sorin M, Pal A, Craig E, Morozov A, et al. (2001) Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1. Nat. Med. 7: 920-926.

80. Christ F, Thys W, De Rijck J, Gijsbers R, Albanese A, et al. (2008) Transportin-SR2 imports HIV into the nucleus. Curr Biol 18(16): 1192-1202.

81. Krishnan L, Matreyek KA, Oztop I, Lee K, Tipper CH, et al. (2010) The requirement for cellular transportin 3 (TNPO3 or TRN-SR2) during infection maps to human immunodeficiency virus type 1 capsid and not integrase. J Virol 84(1): 397-406.

82. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, et al. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J. Biol. Chem. 278: 372-381.

83. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, et al. (2008) Identification of host proteins required for HIV infection through a functional genomic screen. Science 319(5865): 921-926. 10.1126/science.1152725.

84. Konig R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, et al. (2008) Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell 135(1): 49-60.

85. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, et al. (2008) Genome-scale RNAi screen for host factors required for HIV replication. Cell Host Microbe . 10.1016/j.chom.2008.10.004.

86. Pryciak PM, Varmus HE. (1992) Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell 69: 769-780.

87. Fitzgerald ML, Grandgenett DP. (1994) Retroviral integration: In vitro host site selection by avian integrase. J Virol 68(7): 4314-21.

88. Bor Y-, Miller M, Bushman F, Orgel L. (1996) Target sequence preferences of HIV-1 integration complexes in vitro. Virology 222: 238-242.

89. Carteau S, Hoffmann C, Bushman FD. (1998) Chromosome structure and HIV-1 cDNA integration: Centromeric alphoid repeats are a disfavored target. J. Virol. 72: 4005-4014.

90. Pruss D, Bushman FD, Wolffe AP. (1994) Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. Proc. Natl. Acad. Sci. USA 91: 5913-5917.

91. Pruss D, Reeves R, Bushman FD, Wolffe AP. (1994) The influence of DNA and nucleosome structure on integration events directed by HIV integrase. J. Biol. Chem. 269: 25031-25041.

92. Muller H-, Varmus HE. (1994) DNA bending creates favored sites for retroviral

integration: An explanation for preferred insertion sites in nucleosomes. EMBO J. 13: 4704-4714.

93. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110(4): 521-529.

94. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, et al. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol 2(8): E234.

95. Wu X, Li Y, Crise B, Burgess SM. (2003) Transcription start regions in the human genome are favored targets for MLV integration. Science 300(5626): 1749-1751.

96. Barr SD, Leipzig J, Shinn P, Ecker JR, Bushman FD. (2005) Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. J. Virol. 79: 12035-12044.

97. Barr SD, Ciuffi A, Leipzig J, Shinn P, Ecker JR, et al. (2006) HIV integration site selection: Targeting in macrophages and the effects of different routes of viral entry. Mol Ther 14(2): 218-225.

98. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. (2007) HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res 17(8): 1186-1194.

99. Hacker CV, Vink CA, Wardell TW, Lee S, Treasure P, et al. (2006) The integration profile of EIAV-based vectors. Mol Ther 14(4): 536-45.

100. Kang Y, Moressi CJ, Scheetz TE, Xie L, Tran DT, et al. (2006) Integration site choice of a feline immunodeficiency virus vector. J. Virol. 80: 8820-8823.

101. Kim S, Kim N, Dong B, Boren D, Lee SA, et al. (2008) Integration site preference of xenotropic murine leukemia virus-related virus, a new human retrovirus associated with prostate cancer. J Virol 82(20): 9964-9977.

102. Derse D, Crise B, Li Y, Princler G, Lum N, et al. (2007) Human T-cell leukemia virus type 1 integration target sites in the human genome: Comparison with those of other retroviruses. J Virol 81(12): 6731-6741.

103. Faschinger A, Rouault F, Sollner J, Lukas A, Salmons B, et al. (2008) Mouse mammary tumor virus integration site selection in human and mouse genomes. J Virol 82(3): 1360-1367. 10.1128/JVI.02098-07.

104. Berry C, Hannenhalli S, Leipzig J, Bushman FD. (2006) Selection of target sites for mobile DNA integration in the human genome. PLoS Comput Biol 2(11): e157.

105. Rohdewohld H, Weiher H, Reik W, Jaenisch R, Breindl M. (1987) Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J. Virol. 61: 336.

106. Vijaya S, Steffan DL, Robinson HL. (1986) Acceptor sites for retroviral integrations map near DNaseI-hypersensitive sites in chromatin. J. Virol. 60: 683-692.

107. Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, et al. (2006) Retroviral DNA integration: Viral and cellular determinants of target-site selection. PLoS Pathog 2(6): e60.

108. Kirchner J, Connolly CM, Sandmeyer SB. (1995) Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. Science 267(5203): 1488-1491.

109. Zhu Y, Zou S, Wright DA, Voytas DF. (1999) Tagging chromatin with retrotransposons: Target specificity of the sacchromyces Ty5 retrotransposon with the chromosomal localization of Sir3p and Sir4p. Genes Dev. 13: 2738-2749.

110. Bushman F, Miller MD. (1997) Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. J. Virol. 71: 458-464.

111. Tan W, Dong Z, Wilkinson TA, Barbas CF, Chow SA. (2006) Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. J. Virol. 80: 1939-1948.

112. Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, et al. (2005) Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. J Biol Chem 280(27): 25517-23.

113. Ge H, Si Y, Roeder RG. (1998) Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. EMBO J. 17: 6723-6729.

114. Singh DP, Ohguro N, Kikuchi T, Sueno T, Reddy VN, et al. (2000) Lens epithelium-derived growth factor: Effects on growth and survival of lens epithelial cells, keratinocytes, and fibroblasts. Biochem Biophys Res Commun 267(1): 373-381.

115. Singh DP, Fatma N, Kimura A, Chylack LT, Shinohara T. (2001) LEDGF binds to heat shock and stress-related element to activate the expression of stress-related genes. Bio. Biophys. Res. Com. 283: 943-955.

116. Sharma P, Singh DP, Fatma N, Chylack LT,Jr, Shinohara T. (2000) Activation of LEDGF gene by thermal-and oxidative-stresses. Biochem Biophys Res Commun 276(3): 1320-1324.

117. Daniels T, Zhang J, Gutierrez I, Elliot ML, Yamada B, et al. (2005) Antinuclear autoantibodies in prostate cancer: Immunity to LEDGF/p75, a survival protein highly expressed in prostate tumors and cleaved during apoptosis. Prostate 62(1): 14-26.

118. Cohen B, Addadi Y, Sapoznik S, Meir G, Kalchenko V, et al. (2009) Transcriptional regulation of vascular endothelial growth factor C by oxidative and thermal stress is mediated by lens epithelium-derived growth factor/p75. Neoplasia 11(9): 921-933.

119. Ahuja HG, Hong J, Aplan PD, Tcheurekdjian L, Forman SJ, et al. (2000) t(9;11)(p22;p15) in acute myeloid leukemia results in a fusion between NUP98 and the gene encoding transcriptional coactivators p52 and p75-lens epithelium-derived growth factor (LEDGF). Cancer Res 60(22): 6227-6229.

120. Grand FH, Koduru P, Cross NC, Allen SL. (2005) NUP98-LEDGF fusion and t(9;11) in transformed chronic myeloid leukemia. Leuk Res 29(12): 1469-1472.

121. Sutherland HG, Newton K, Brownstein DG, Holmes MC, Kress C, et al. (2006) Disruption of Ledgf/Psip1 results in perinatal mortality and homeotic skeletal transformations. Mol Cell Biol 26(19): 7201-7210.

122. Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, et al. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev 21(14): 1767-1778.

123. Vanegas M, Llano M, Delgado S, Thompson D, Peretz M, et al. (2005) Identification of the LEDGF/p75 HIV-1 integrase interaction domain and NLS reveals NLS-independent chromatin tethering. J. Cell. Sci. 118: 1733-1743.

124. Turlure F, Maertens G, Rahman S, Cherepanov P, Engelman A. (2006) A

tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. Nucleic Acids Res 34(5): 1653-1675.

125. Llano M, Vanegas M, Hutchins N, Thompson D, Delgado S, et al. (2006) Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. J Mol Biol 360(4): 760-773.

126. Stec I, Nagl SB, van Ommen GJ, den Dunnen JT. (2000) The PWWP domain: A potential protein-protein interaction domain in nuclear proteins influencing differentiation? FEBS Lett 473(1): 1-5.

127. Qiu C, Sawada K, Zhang X, Cheng X. (2002) The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Biol 9(3): 217-224.

128. Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, et al. (2003) The tudor domain 'royal family': Tudor, plant agenet, chromo, PWWP and MBT domains. Trends Biochem Sci 28(2): 69-74.

129. Llano M, Vanegas M, Fregoso O, Saenz D, Chung S, et al. (2004) LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. J. Virol. 78: 9524-9537.

130. Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, et al. (2005) The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. J Biol Chem 280(18): 17841-17847.

131. Cherepanov P, Devroe E, Silver PA, Engelman A. (2004) Identification of an evolutionarily-conserved domain of LEDGF/p75 that bind HIV-1 integrase. J. Biol. Chem. 279: 48883-48892.

132. Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, et al. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J. Biol. Chem. 278: 33528-33539.

133. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A. (2005) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc Natl Acad Sci U S A 102: 17308-17313.

134. Maertens GN, Cherepanov P, Engelman A. (2006) Transcriptional co-activator p75 binds and tethers the myc-interacting protein JPO2 to chromatin. J Cell Sci 119(Pt 12): 2563-71.

135. Bartholomeeusen K, Christ F, Hendrix J, Rain JC, Emiliani S, et al. (2009) Lens epithelium-derived growth factor/p75 interacts with the transposase-derived DDE domain of PogZ. J Biol Chem 284(17): 11467-11477.

136. Yokoyama A, Cleary ML. (2008) Menin critically links MLL proteins with LEDGF on cancer-associated target genes. Cancer Cell 14(1): 36-46.

137. Cherepanov P, Sun ZY, Rahman S, Maertens G, Wagner G, et al. (2005) Solution structure of the HIV integrase-binding domain in LEDGF/p75. Nat. Struct. Mol. Biol. 12: 526-532.

138. De Rijck J, Vandekerckhove L, Gijsbers R, Hombrouck A, Hendrix J, et al. (2006) Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. J Virol 80(23): 11498-11509.

139. Llano M, Delgado S, Vanegas M, Poeschla EM. (2004) LEDGF/p75 prevents proteasomal degradation of HIV-1 integrase. J. Biol. Chem. 279: 55570-55577.

140. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, et al. (2006) An essential role for LEDGF/p75 in HIV integration. Science .

141. Vandegraaff N, Devroe E, Turlure F, Silver PA, Engelman A. (2006) Biochemical and genetic analyses of integrase-interacting protein lens epithelium-derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication. Virology 346: 415-426.

142. Vandekerckhove L, Christ F, Van Maele B, De Rijck J, Gijsbers R, et al. (2006) Transient and stable knockdown of the integrase cofactor LEDGF/p75 reveals its role in the replication cycle of human immunodeficiency virus. J Virol 80(4): 1886-96.

143. Zielske SP, Stevenson M. (2006) Modest but reproducible inhibition of human immunodeficiency virus type 1 infection in macrophages following LEDGFp75 silencing. J Virol 80(14): 7275-7280.

144. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, et al. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med 11(12): 1287-1289.

145. Li L, Olvera JM, Yoder K, Mitchell RS, Butler SL, et al. (2001) Role of the non-homologous DNA end joining pathway in retroviral infection. EMBO J. 20: 3272-3281.

146. Kilzer JM, Stracker TH, Beitzel B, Meek K, Weitzman MD, et al. (2003) Roles of host cell factors in circularization of retroviral DNA. Virology 314: 460-467.

147. Delenda C. (2004) Lentiviral vectors: Optimization of packaging, transduction and gene expression. J Gene Med 6 Suppl 1: S125-38.

148. Sinn PL, Sauter SL, McCray PB,Jr. (2005) Gene therapy progress and prospects: Development of improved lentiviral and retroviral vectors--design, biosafety, and production. Gene Ther 12(14): 1089-1098.

149. Shimotohno K, Temin HM. (1981) Formation of infectious progeny virus after insertion of herpes simplex thymidine kinase gene into DNA of an avian retrovirus. Cell 26(1 Pt 1): 67-77.

150. Mann R, Mulligan RC, Baltimore D. (1983) Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. Cell 33: 153-159.

151. Verma IM, Somia N. (1997) Gene therapy-promises, problems and prospects. Nature 389: 239-242.

152. Edelstein M. (2009) Gene therapy clinical trials worldwide. .

153. Logan AC, Nightingale SJ, Haas DL, Cho GJ, Pepper KA, et al. (2004) Factors influencing the titer and infectivity of lentiviral vectors. Hum Gene Ther 15(10): 976-988.

154. Zufferey R, Donello JE, Trono D, Hope TJ. (1999) Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. J Virol 73(4): 2886-92.

155. Krall WJ, Skelton DC, Yu XJ, Riviere I, Lehn P, et al. (1996) Increased levels of spliced RNA account for augmented expression from the MFG retroviral vector in hematopoietic cells. Gene Ther 3(1): 37-48.

156. Cronin J, Zhang XY, Reiser J. (2005) Altering the tropism of lentiviral vectors through pseudotyping. Curr Gene Ther 5(4): 387-398.

157. Dull T, Zufferey R, Kelly M, Mandel RJ, Nguyen M, et al. (1998) A third-generation lentivirus vector with a conditional packaging system. J Virol 72(11): 8463-71.

158. V E, Bang O. (1908) Experimentelle leukamie bei huhnern. Zentralbl. Bakteriol.

Parasitenkd. Infectionskr. Hyg. Abt. Orig. 46: 595-609.

159. Rous P. (1911) A sarcoma of the fowl transmissible by an agent seperable from the tumor cells. J. Exp. Med. 13: 397-411.

160. Engelke U, Lipsick JS. (1994) Transformation of myelomonocytic cells by the avian myeloblastosis virus is determined by the v-myb oncogene, not by the unique long terminal repeats of the virus. J Virol 68(4): 2752-2755.

161. Kirsten WH, Mayer LA. (1967) Morphologic responses to a murine erythroblastosis virus. J Natl Cancer Inst 39(2): 311-335.

162. Wolff L, Ruscetti S. (1985) Malignant transformation of erythroid cells in vivo by introduction of a nonreplicating retrovirus vector. Science 228(4707): 1549-1552.

163. Deichmann A, Hacein-Bey-Abina S, Schmidt M, Garrigue A, Brugman MH, et al. (2007) Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. J Clin Invest 117(8): 2225-2232.

164. Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, et al. (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest 118(9): 3132-3142.

165. Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C, Hubank M, et al. (2008) Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. J Clin Invest 118(9): 3143-3150.

166. Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, et al. (2006) Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat Med 12(4): 401-409.

167. Stein S, Ott MG, Schultze-Strasser S, Jauch A, Burwinkel B, et al. (2010) Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nat Med .

168. Baum C, Kustikova O, Modlich U, Li Z, Fehse B. (2006) Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. Hum Gene Ther 17(3): 253-263.

169. Recchia A, Bonini C, Magnani Z, Urbinati F, Sartori D, et al. (2006) Retroviral vector integration deregulates gene expression but has no consequence on the

biology and function of transplanted T cells. Proc Natl Acad Sci U S A 103(5): 1457-1462.

170. Kustikova OS, Schiedlmeier B, Brugman MH, Stahlhut M, Bartels S, et al. (2009) Cell-intrinsic and vector-related properties cooperate to determine the incidence and consequences of insertional mutagenesis. Mol Ther 17(9): 1537-1547.

171. Hacein-Bey H, Cavazzana-Calvo M, Le Deist F, Dautry-Varsat A, Hivroz C, et al. (1996) Gamma-c gene transfer into SCID X1 patients' B-cell lines restores normal high-affinity interleukin-2 receptor expression and function. Blood 87(8): 3108-3116.

172. Zychlinski D, Schambach A, Modlich U, Maetzig T, Meyer J, et al. (2008) Physiological promoters reduce the genotoxic risk of integrating gene vectors. Mol Ther 16(4): 718-725.

173. Ramezani A, Hawley TS, Hawley RG. (2008) Combinatorial incorporation of enhancer-blocking components of the chicken beta-globin 5'HS4 and human T-cell receptor alpha/delta BEAD-1 insulators in self-inactivating retroviral vectors reduces their genotoxic potential. Stem Cells 26(12): 3257-3266.

174. Ramezani A, Hawley TS, Hawley RG. (2008) Reducing the genotoxic potential of retroviral vectors. Methods Mol Biol 434: 183-203.

175. Miyoshi H, Blomer U, Takahashi M, Gage FH, Verma IM. (1998) Development of a self-inactivating lentivirus vector. J Virol 72(10): 8150-8157.

176. Zufferey, R., Dull, T., Mandel, R., Bukovsky, A., Quiroz, D., Naldini, L., Trono,D. (1998) Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. J. Virol. 72(12): 9873-80.

177. Modlich U, Bohne J, Schmidt M, von Kalle C, Knoss S, et al. (2006) Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity. Blood 108(8): 2545-2553.

178. Modlich U, Schambach A, Brugman MH, Wicke DC, Knoess S, et al. (2008) Leukemia induction after a single retroviral vector insertion in Evi1 or Prdm16. Leukemia 22(8): 1519-1528.

179. Iwakuma T, Cui Y, Chang LJ. (1999) Self-inactivating lentiviral vectors with U3 and U5 modifications. Virology 261(1): 120-132.

180. Raineri I, Senn HP. (1992) HIV-1 promotor insertion revealed by selective

detection of chimeric provirus-host gene transcripts. Nucleic Acids Res 20(23): 6261-6266.

181. Trono D. (2003) Virology. picking the right spot. Science 300(5626): 1670-1671.

182. De Palma M, Montini E, Santoni de Sio FR, Benedicenti F, Gentile A, et al. (2005) Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoetic cells. Blood 105: 2307-2315.

183. Montini E, Cesana D, Schmidt M, Sanvito F, Ponzoni M, et al. (2006) Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. Nat Biotechnol 24(6): 687-696. 10.1038/nbt1216.

184. Montini E, Cesana D, Schmidt M, Sanvito F, Bartholomae CC, et al. (2009) The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. J Clin Invest 119(4): 964-975.

185. Modlich U, Navarro S, Zychlinski D, Maetzig T, Knoess S, et al. (2009) Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. Mol Ther 17(11): 1919-1928.

186. Levine BL, Humeau LM, Boyer J, MacGregor RR, Rebello T, et al. (2006) Gene transfer in humans using a conditionally replicating lentiviral vector. Proc Natl Acad Sci U S A 103(46): 17372-17377.

187. Wang GP, Levine BL, Binder GK, Berry CC, Malani N, et al. (2009) Analysis of lentiviral vector integration in HIV+ study subjects receiving autologous infusions of gene modified CD4+ T cells. Mol Ther .

188. Cartier N, FAU - Hacein-Bey-Abina S, Hacein-Bey-Abina S, FAU - Bartholomae CC, Bartholomae CC, et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. - Science.2009 Nov 6;326(5954):818-23. (1095-9203 (Electronic)).

189. Ciuffi A, Bushman FD. (2006) Retroviral DNA integration: HIV and the role of LEDGF/p75. Trends Genet. Epub.

190. Gijsbers R, Ronen K, Vets S, Malani N, De Rijck J, et al. (2010) LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. Mol Ther. advance online publication Jan 5, 2010.

CHAPTER 2 – THE ROLE OF PSIP1/LEDGF/p75 IN LENTIVIRAL INFECTIVITY AND INTEGRATION TARGETING

The contents of this chapter have been published as part of:
Marshall HM*, Ronen K*, Berry C, Llano M, Sutherland H, Saenz D, Bickmore W, Poeschla E, Bushman FD. (2007) Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. PLoS One. 2(12):e1340.
*equal contribution

2.1 Abstract

To replicate, lentiviruses such as HIV must integrate DNA copies of their RNA genomes into host cell chromosomes. Lentiviral integration is favored in active transcription units, which allows efficient viral gene expression after integration, but the mechanisms directing integration targeting are incompletely understood. A cellular protein, PSIP1/LEDGF/p75, binds tightly to the lentiviral-encoded integrase protein (IN), and has been reported to be important for HIV infectivity and integration targeting. Here we report studies of lentiviral integration targeting in murine cells with homozygous gene trap mutations in the *PSIP1/LEDGF/p75* locus. Infections with vectors derived from HIV and equine infections anemia virus (EIAV) were compared. Integration acceptor sites were analyzed by DNA bar coding and pyrosequencing. In PSIP1/LEDGF/p75-depleted murine embryonic firbroblasts, reductions were seen in lentiviral infectivity compared to controls. Reductions in integration in transcription units were seen, paralleling studies of human models and a different mutant mouse line. Integration did not become random, however – integration in transcription units was still favored, though to a reduced degree. New

trends also appeared, including favored integration near CpG islands. In addition, we carried out a bioinformatic study of 15 HIV integration site data sets in different cell types, which showed that the frequency of integration in transcription units was correlated with the cell-type specific levels of *PSIP1/LEDGF/p75* expression.

2.2 Introduction

Early steps of retroviral replication involve reverse transcription to generate a DNA copy of the viral RNA genome, and integration, which results in the covalent connection of the viral DNA to host cell DNA (for reviews see [1, 2]). The question of where retroviruses target DNA integration is central to understanding viral host interactions. For the virus, selection of favorable sites for viral DNA integration assists efficient expression of the viral genome after integration [3-6]. For the host, viral DNA integration can either activate or inactivate gene transcription. One consequence of integration can be insertional activation of oncogenes and transformation to malignant growth [1, 2, 7, 8]. Here we present data on the role of a host-cell encoded protein, PSIP1/LEDGF/p75, that guides integration site selection by lentiviruses, the viral genus including HIV (henceforth we use "LEDGF/p75" because this name is widely used in the HIV field).

LEDGF/p75 first came to the attention of the retrovirus field when it was identified in affinity-based screens for its tight binding to HIV IN [9-11]. LEDGF/p75 tethers ectopically-expressed HIV IN to chromatin [9, 10, 12, 13], through specific binding domains [14-17], and also protects IN from proteasomal degradation [18]. LEDGF/p75 binding is specific for lentiviral IN proteins (e. g.

those of HIV, SIV, FIV, and EIAV) [12, 19, 20], which makes it appealing as a candidate tethering factor since all lentiviruses tested (HIV, SIV, FIV, and EIAV) show favored integration in active transcription units [5, 21-32]. The crystal structure of the catalytic domain of HIV IN (residues 50-212) bound to the integrase binding domain (IBD) was solved, which showed that a pair of LEDGF/p75-IBD molecules could bind at symmetry-related positions at the interface of the IN catalytic domain dimer [33, 34].

Early attempts to determine whether LEDGF/p75 was important for efficient HIV replication used RNAi knockdowns in human cells, which had either no effect or quantitatively modest effects on infection [12, 13, 35, 36]. This now appears to be because incomplete knockdowns left biologically significant amounts of protein present. More recently, human SupT1 cells with intensified RNAi knockdowns showed infectivity drops of 30-fold by either HIV or another lentivirus, feline immunodeficiency virus (FIV), and combining this with dominant interfering proteins derived from the LEDGF/p75-IBD produced 560-fold inhibition of infection [37]. These findings are supported by additional studies in human cell lines [35, 38, 39].

Early knockdowns of LEDGF/p75 were also analyzed for effects on targeting of HIV integration [40]. Knockdowns in three cell types were studied, and in each integration frequency within transcription units was reduced. In addition, other effects were seen, including an increase in the content of G/C bases around sites of HIV integration in the knockdown cells. These data supported the idea that LEDGF/p75 acted as a tethering factor, binding to both HIV and chromatin to direct

HIV integration into active genes. In support of the tethering model, artificial fusion proteins in which the LEDGF/p75 IBD was fused to the sequence specific DNA binding domain of phage lambda repressor were shown to direct favored integration i*n vitro* near repressor binding sites [24]. Also supporting the tethering idea, function of LEDGF/p75 in promoting HIV replication requires that both ends of the putative LEDGF/p75 tether be intact [37].

However, key questions still remained on the role of LEDGF/p75. In all the models studied, HIV continued to favor integration within active transcription units. This could either be because residual LEDGF/p75 remaining in the knockdown was sufficient for residual targeting activity, or because additional host cell factors also contribute independently to targeting HIV integration. In an effort to address this issue, we studied mouse cells containing homozygous gene trap mutations at the *LEDGF/p75* locus developed by Sutherland and coworkers [42]. Vectors derived from equine infectious anemia virus (EIAV) were used in many of the experiments, allowing effects on HIV and EIAV to be compared. Studies of both lentiviruses provided strong evidence for the role of LEDGF/p75 in promoting efficient infection and targeting integration in transcription units. Additionally, new integration preferences emerged in LEDGF/p75-depleted cells, and some targeting to transcription units persisted. In data not shown, Heather Marshall in the lab conducted similar experiments with the human SupT1 T-cell line with intensified RNAi developed by Llano et al. [37], and obtained similar results, confirming that murine cells are an appropriate model for LEDGF/p75 function.

While our study was under way, Shun *et al.* prepared a mouse strain in which part of the *LEDGF/p75* locus was flanked by Cre recombination sites [41], and the *LEDGF/p75* exon was deleted by exposure to Cre recombinase. Mouse embryonic fibroblasts were then studied for effects on infection with HIV reporter viruses. In agreement with our studies, these cells showed a 20-fold reduction in infectivity by HIV, and also a reduction in integration frequency in transcription units that was stronger than that reported in human cell knockdowns by Ciuffi *et al.* [40]. The mouse cells also showed some new targeting features in the LEDGF/p75-depleted cells, including increased integration near CpG islands.

In addition to these data on manipulated cell models, we also present additional bioinformatic studies of 15 published HIV integration site data sets in different cell types, which revealed a strong correlation between cell type specific *LEDGF/p75* expression levels and the proportion of HIV integration sites in transcription units. These data provide further support for the generality of LEDGF/p75 as a determinant of integration target site selection for lentiviruses, including in primary cells where LEDGF/p75 levels were not artificially reduced.

2.3 Materials and Methods

Cell lines

MEFs were extracted from wild-type and knockout embryos at 13.5 dpc [60] and cultured in DMEM with 10% FBS, 50μg/ml gentamycin, 110μM beta-mercaptoethanol, 1X non-essential amino acids, 100μM sodium pyruvate. Primary

MEFs (prMEFs) were immortalized by the 3T3 protocol, by splitting cells every 3 days to a density of $6X10^4$ cells/ml [61].

Viral particle production and infections

VSV-G pseudotyped HIV vector particles were produced by Lipofectamine transfection of 293T cells with p156RRLsin-PPTCMVGFPWPRE [62], the packaging construct pCMVdeltaR9 [63], and the vesicular stomatitis virus G-producing pMD.G construct. EIAV vector particles were likewise produced by transfection with p6.1G3CeGFPw (M. Patel and J. Olsen, University of North Carolina, Chapel Hill unpublished), the packaging construct pEV53B [64], and the vesicular stomatitis virus G-producing plasmid pVSVG into 293T cells. Viral supernatant was harvested 38 hours after transfection, filtered through 0.22µm filters, concentrated by filtration through a Centricon, treated with DNase I, and stored frozen at -80°C. HIV titer was quantified by p24 ELISA.

For HIV infection, cells were plated onto 6-well plates at a density of $3X10^5$ cells per well and each well infected with 1µg p24. For EIAV, cells were plated into 24-well plates at a density of $4X10^4$ cells per well, and each well infected with 100µl concentrated virus. Infections were performed overnight in the presence of 10µg/ml DEAE-dextran. 10 independent HIV infections and 5 EIAV infections were performed per genotype. 48 hours after infection, 90% of cells were harvested for integration site cloning and the remainder passaged for an additional 2 weeks to dilute unintegrated products of reverse transcription and used for QPCR analysis of integration efficiency.

Infectivity tests

For quantitative PCR analysis, infected cells were passaged for 2 weeks following infection to dilute unintegrated products of reverse transcription, then genomic DNA was extracted using the Qiagen DNeasy tissue extraction kit. QPCR using HIV late-RT primers and probe was carried out as described in [44] using 50ng genomic DNA as template. For EIAV, primer and probe sequences are described in Table S1. 25ng of SupT1 genomic DNA was used as template, 50ng of MEF genomic DNA. QPCR was performed using Applied Biosystems 2X FAST universal master mix and Applied Biosystems FAST PCR machine.

Integration site amplification

Integration sites were isolated and sequenced by ligation-mediated PCR essentially as described previously [46]. Genomic DNA was extracted from infected cells using the Qiagen DNeasy tissue extraction kit. Up to 2µg of DNA from each infection was digested overnight using MseI. This was followed by digestion to prevent amplification of internal viral fragments (from the 5' LTR) and plasmid backbone with SacI and DpnI in the case of HIV, and XmaI and DpnI in the case of EIAV. Linkers were then ligated onto digested products (oligonucleotide sequences listed below) and nested PCR performed from ligation products. Nested PCR primers contained 4 or 8 nt barcode sequences between the sequencing primer and LTR-binding portions. These enabled pooling of all PCR products into one sequencing reaction and subsequent separation of sequences by decoding the barcodes. Amplification products were gel-purified and sent to the Interdisciplinary Center for

Biotechnology Research at the University of Florida and the Virginia Bioinformatics Institute Core Laboratory Facility for pyrosequencing. Sequences have been deposited in the NCBI database, under accession numbers GS773309-GS815944.

Oligos used

Primer name	Primer Sequence (5' ==> 3')
EIAV Primer PCR1	CCTGTCTCTAGTTTGTCTGTTCG
EIAV nested PCR Forward	gccttgccagcccgctcagxxxxAGTTTGTCTGTTCGAGATCCTACA*
HIV Primer PCR1	CTTAAGCCTCAATAAAGCTTGCCTTGAG
HIV nested PCR Forward	gccttgccagcccgctcagxxxxAGACCCTTTTAGTCAGTGTGGAAAATC**
MseI Linker Primer for nested PCR	gcctccctcgcgccatcagAGGGCTCCGCTTAAGGGAC***
MseI Linker Primer for PCR 1	GTAATACGACTCACTATAGGGC
MseI linker positive strand	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC
MseI linker negative strand	[Phosp]-TAGTCCCTTAAGCGGAG-[AmC7-Q]****
EIAV Q-PCR forward	CCG CAA TAA CCA CAT TTG TGA CGC
EIAV Q-PCR reverse	GCA GAA TCT GAG TGC CCA ATT GTC AG
EIAV Q-PCR probe	FAM-AGT TCC GCA TTG GTG ACG CGT TAA GT-Black_Hole_Quencher
HIV Q-PCR forward	TGTGTGCCCGTCTGTTGTGT
HIV Q-PCR reverse	GAGTCCTGCGTCGAGAGAGC
HIV Q-PCR probe	FAM-CAGTGGCGCCCGAACAGGGA-TAMRA

*__primer B__ - barcode - EIAV 3'LTR PRIMER (nested PCR)
**__primer B__ - barcode - HIV 3'LTR PRIMER (nested PCR)
***__primer A__-MSEI LINKER PRIMER (for nested PCR)
**** modifications- [Phosp]=5' phosphate, [AmC7-Q]=3' amine

Bioinformatic analysis.

Integration sites were judged to be authentic if the sequences had a best unique hit when aligned to the murine (mm8 draft) using BLAT, and the alignment began within 3bp of the viral LTR end and had >98% sequence identity. Detailed statistical methods are described in [55].

To control for possible biases in isolating integration sites due to restriction enzyme sequence distribution, three or ten matched random controls were computationally generated for each experimental integration site that were the same distance from the closest MseI restriction site as the experimental site.

Integration site counts in various genomic annotations were compared with matched random controls by the Fisher's exact test. Additionally, multiple regression

models for integration intensity were applied, as described in [55].

For analysis of correlations with gene activity in murine integration sites (Figures 2-3 and 2-4), transcriptional profiling data from wild-type MEFs analyzed on the MGU74Av2 Affymetrix microarray were used. Genes represented on the microarray were ranked by expression level and divided into 4 bins based on expression level. Integration sites found within genes in each bin were counted as a proportion of sites found within genes in all bins.

For the analysis of relative gene activity in Figure 2-5, data from two types of Affymetrix chips were used (HU95A and HU133A). Two probe sets querying *LEDGF/p75* but not *p52* were available on each chip (For HU95: 39243_s_at and 37622_r_at; for HU133: 209337_at and 205961_s_at). To account for differences in the sensitivities arising from the different chip designs and probe sets, the values for each cell type were first ranked for each probe set and chip combination, then the ranked values pooled in the final data set.

2.4 Results

Efficiency of lentivirus infection in murine cells disrupted at *LEDGF/p75*

Human cells stringently depleted for LEDGF/p75 [37] and murine embryonic fribroblasts (MEFs) from knockout mice [41] show a block in HIV infection. We wanted to verify this block in lentiviral infection in MEFs containing the gene trap disruption of *LEDGF/p75* reported by Sutherland and colleagues [42]. Because residual expression is sometimes detected in gene trap alleles, we used quantitative RT-PCR to determine the fraction of *LEDGF/p75* messages disrupted by the gene

trap insertion. In samples from homozygous mutant (-/-) cells, amplification of correct *LEDGF/p75* message was sporadically detected at high PCR cycle numbers, suggesting that rare correctly spliced messages were formed. However, quantification of correct message formation using SyberGreen quantitative PCR showed expression of *LEDGF/p75* to be below the limit of detection in the -/- cells, corresponding to a reduction of at least 32-fold compared to the wild type (+/+) cells (data not shown). Sutherland and coworkers reported LEDGF/p75 protein to be undetectable [42].

We analyzed infection of MEFs isolated from embryos of +/+ and homozygous mutant -/- mice after infection with HIV and EIAV. Like HIV IN, EIAV IN is known to bind LEDGF/p75 [20], and EIAV is also known to integrate in active transcription units [31], so EIAV is a suitable model for analysis of the influence of LEDGF/p75 on lentivirus infection. Integration was measured by infecting cells, maintaining the cells in culture for two weeks to allow loss of unintegrated DNA [44], then quantifying the viral DNA by TaqMan PCR. HIV integration was reduced ~five fold in the *LEDGF/p75* -/- MEFs (Figure 2-1), and EIAV integration was reduced >50 fold. Thus in the presence of a homozygous mutation of *LEDGF/p75*, lentiviral integration was strongly reduced but not eliminated.

DNA bar coding and pyrosequencing to analyze integration site placement

We used the pyrosequencing technology commercialized by 454 Life Sciences [45] to sequence genomic DNA flanking integrated proviruses. Briefly, genomic DNA was isolated and cleaved with restriction enzymes. DNA linkers were

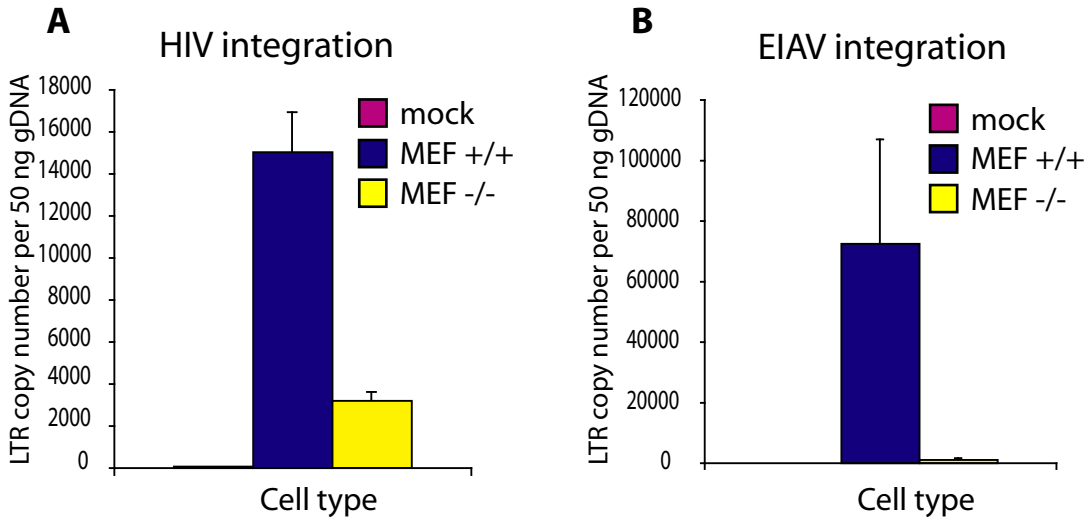

Figure 2-1. Efficiency of lentiviral infection in control and LEDGF/p75-disrupted cells. Integration of HIV (**A**) and EIAV (**B**) was measured by quantitative PCR. +/+ control; -/- homozygous LEDGF/p75-disrupted.

ligated onto the cleaved ends, then host-virus DNA junctions were amplified using one primer complementary to the linker and one complementary to the viral DNA end. A second round of PCR was used to improve specificity and to add recognition sites for the 454 primers necessary for the emulsion PCR step preceding pyrosequencing [46]. Pooled DNAs were then sequenced.

Use of DNA bar coding allowed multiple integration site populations to be studied in parallel [47-49]. The viral DNA primer used in the second round of amplification contained a short recognition sequence (4-8 bases) abutting the 454 primer that was different for each sample tested. These 4-8 bases are the first determined in pyrosequencing reads. Thus use of bar coding allowed many samples to be pooled for sequence determination, then the reads could be sorted into individual experiments by bar code. A total of 1757 unique integration site sequences from different virus and cell combinations were determined using this method (Table 2-1). We analyzed integration sites in murine embryonic fibroblasts (MEFs) derived from the *LEDGF/p75* homozygous gene trap (-/-) and control (+/+) mice [42] after infection with HIV and EIAV. Cells that had been immortalized in culture (iMEF) were compared to primary MEFs (prMEFs). For all the features discussed below the results were identical for iMEFs and prMEFs (data not shown), so the two data sets were pooled in what follows.

Consensus sequences at lentiviral integration sites in murine cells disrupted at *LEDGF/p75*

Cell line	Description and LEDGF/p75 status	Virus	Number of Integration Sites	Source of sequences analyzed
iMEF +/+	Murine embryonic fibroblasts from wild-type mice (immortalized)	HIV vector	574	This report
iMEF -/-	Murine embryonic fibroblasts from gene-trap mice (immortalized)	HIV vector	287	This report
prMEF +/+	Murine embryonic fibroblasts from wild-type mice (primary)	HIV vector	531	This report
prMEF -/-	Murine embryonic fibroblasts from gene-trap mice (primary)	HIV vector	209	This report
iMEF +/+	Murine embryonic fibroblasts from wild-type mice (immortalized)	EIAV vector	70	This report
iMEF -/-	Murine embryonic fibroblasts from gene-trap mice (immortalized)	EIAV vector	86	This report

Table 2-1. Integration site data sets used in this study.

As a first step in the analysis, the favored target DNA sequences at the point of integration were compared in the presence and absence of LEDGF/p75. Alignment of target DNA sequences at integration sites has revealed weak inverted repeat consensus sequences [50-55], the symmetry arising because the favored sequence features at each of the two viral DNA ends are the same. The presence of this consensus sequence can be a strong predictor of integration targeting specificity, particularly over short intervals [55]. For HIV, the favored consensus sequence, TDG↓GTWACCHA (where the arrow represents the site of integration) has been synthesized and shown to be a favored integration target site for HIV preintegration complexes *in vitro* [52]. EIAV has been reported to favor integration in an A/T rich palindromic consensus sequence [31]

Integration site sequences were aligned to determine the consensus palindromic sequence at the point of integration, and results were compared for the +/+ and -/- MEFs for each virus (Figure 2-2). In both cases, integration in the +/+ MEFs showed the weak consensus seen previously for HIV and EIAV. No major differences were seen in the -/- MEFs, consistent with previous reports of LEDGF/p75 depletion [40, 41].

HIV integration targeting in murine cells disrupted at *LEDGF/p75*

Genome-wide studies of HIV integration targeting in murine cells are presented in this section and analysis of EIAV integration in murine cells is described in the next section. The data are summarized in Table 2-2 and Figures 2-3 and 2-4.

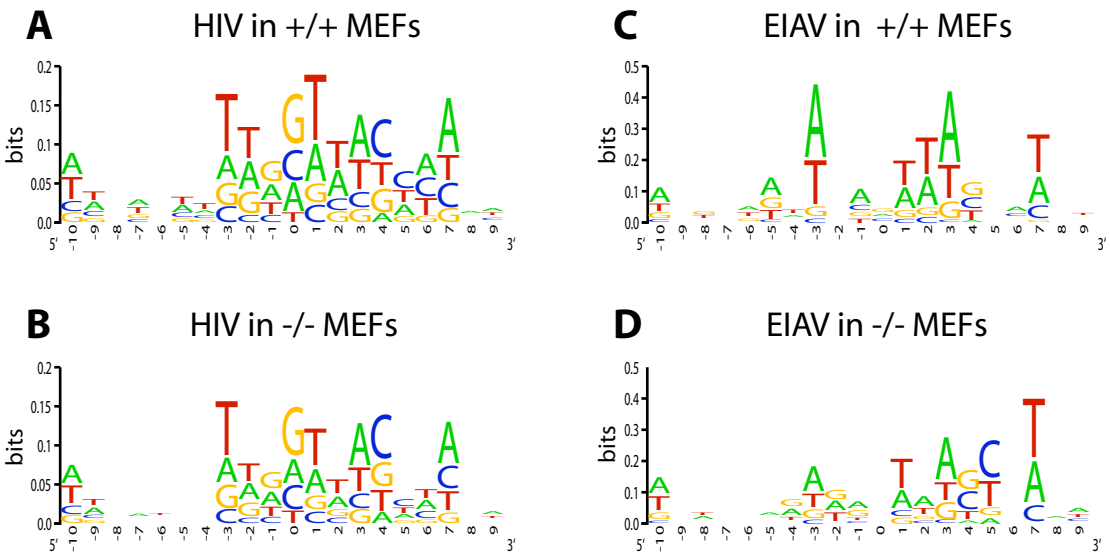

Figure 2-2. Integration site consensus sequence for lentiviral infection of murine control and LEDGF/p75-disrupted cells. A. HIV in +/+ MEFs. **B.** HIV integration in -/- MEFs. **C.** EIAV integration in +/+ MEFs. **D.** EIAV integration in -/- MEFs. The diagrams were generated using the WebLOGO program (weblogo.berkeley.edu). The y-axis indicates bits of information – perfect conservation of a base would score as two bits.

HIV integration in transcription units was decreased in the -/- *LEDGF/p75* gene trap cells compared with wild-type. In wild-type cells, 54.3% of experimental integration sites were in RefSeq genes (see Table 2-2), a significant enrichment over the 28% seen in the matched random controls (see Figure 2-3). In -/- MEFs, 38.7% of sites were in RefSeq transcription units, a value that is significantly less than in the +/+ MEFs ($p<0.0001$ by the Fisher's exact test). Significant differences were seen when the analysis was repeated using other gene catalogs as well (Table 2-2).

We also analyzed the proximity of HIV integration sites to CpG islands. In wild-type cells integration within 2kb of CpG islands was significantly disfavored compared with random, while in knockout cells integration was 10-fold enriched over random ($P<0.0001$ for the comparison between genotypes)

The frequency of integration within 5kb of RefSeq gene 5' ends showed a similar pattern (Table 2-2). Integration levels around gene 5' ends were significantly higher than random in the +/+ cells (10.9% of sites), but in the knockout a further increase was observed (15.5% of sites within 5kb of gene 5' end) achieving $P=0.014$ for the comparison between cell types (Fisher's exact test).

Knockdown of LEDGF/p75 has previously been shown to result in an increase in the G/C content of HIV integration site sequences [40]. We therefore analyzed the frequency of integration in regions of varying G/C content (Figure 2-3), revealing that integration was significantly increased in more G/C rich regions in the -/- MEFs ($P=4e-16$).

Data Set	Frequency in Genomic Feature (%)				
	Transcription Units			<2kb CpG Island	<5kb Gene 5' End
	Known	RefSeq	Ensemble		
HIV in +/+ MEF	58.6***	54.3***	60.7***	0.7*	10.9***
HIV in -/- MEF	42.9***	38.7***	46.0***	6.5***	15.5***
EIAV in +/+ MEF	62.9***	58.6***	64.3***	1.4	5.7
EIAV in -/- MEF	41.9	38.4	45.3	12.8***	25.6***
Random Control	29.7	28	32	1.7	6.8

Table 2-2. Integration frequency in the presence and absence of LEDGF/p75 near mapped genomic features in the murine genome. Significant deviation from matched random controls according to the Fisher's exact test is denoted by * (***p<0.0001, **p<0.01, *p<0.05). The 'random control' set shown is the matched random control set for the 'HIV in +/+ MEF' integration set (see Materials and Methods for generation of matched random controls).

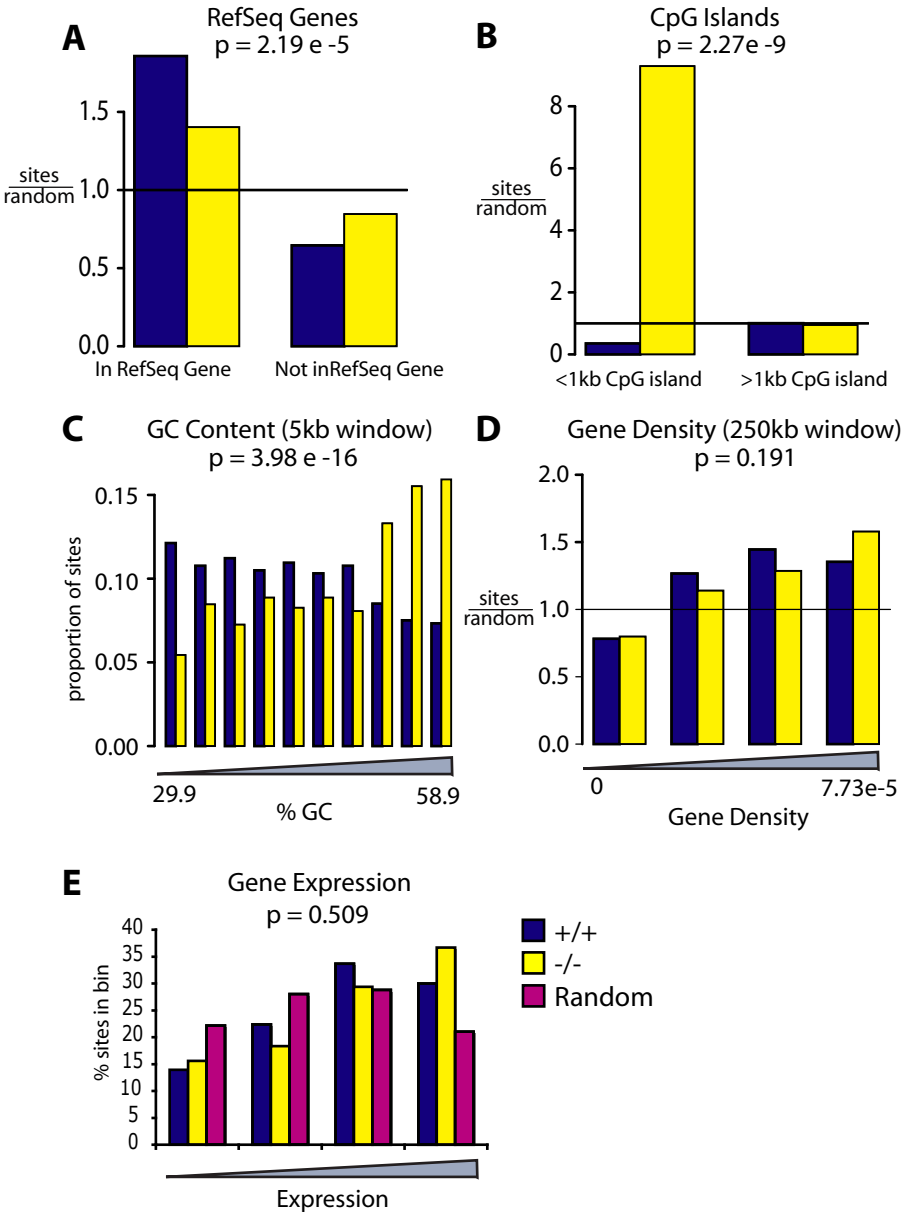

Figure 2-3. HIV integration distributions in control and LEDGF/p75-disrupted cells. Integration site distributions are shown relative to genomic features. **A.** RefSeq genes. **B.** CpG islands (plus or minus 1 kb). **C.** G/C content. Integration sites from unmodified and knockdown cells were pooled and divided into 10 equal bins of increasing GC content, and sites in each cell type plotted for each bin. **D.** Gene density (250kb window). **E.** Relative gene expression intensity. For each value in A-B and D, the measured value for the integration site population was divided by that of the matched random control to emphasize the departure of the experimental data from random. P values shown are based on regression analysis (A-C) or Chi Square test for trend (D-E).

A variety of features analyzed did not show significant differences between genotypes, including the response to gene density (Figure 2-3D) and the relationship between gene activity and integration frequency (Figure 2-3E). We return to the implications of these findings in the Discussion.

EIAV integration targeting in murine cells disrupted at *LEDGF/p75*

Consistent with previous reports [31], EIAV's integration distribution was similar to HIV's. Likewise, the effect of LEDGF/p75 was similar. Integration in transcription units was decreased in the -/- *LEDGF/p75* gene trap cells compared with wild-type, from 58.6% of sites (see Table 2-2 and Figure 2-4) to 38.4% (p=0.016 for the comparison between genotypes by the Fisher's exact test).

We also analyzed the proximity of EIAV integration sites to CpG islands and gene 5' ends. In wild-type cells integration within 2kb of CpG islands was not significantly different from random, while in knockout cells integration was 13-fold enriched over random (P=0.0086; Fisher's exact test). Similarly, integration levels around gene 5' ends were not significantly different from random in the +/+ cells (5.7% of sites), whereas in the knockout a significant enrichment was observed (25.6% of sites) achieving P=0.014 for the comparison between cell types (Fisher's exact test).

We analyzed the correlation between integration frequency and G/C content using a 5kb window around the integration site. A significant difference between genotypes was found, with sites from -/- cells being found in more G/C-rich regions (P=0.001, using regression analysis, Figure 2-4).

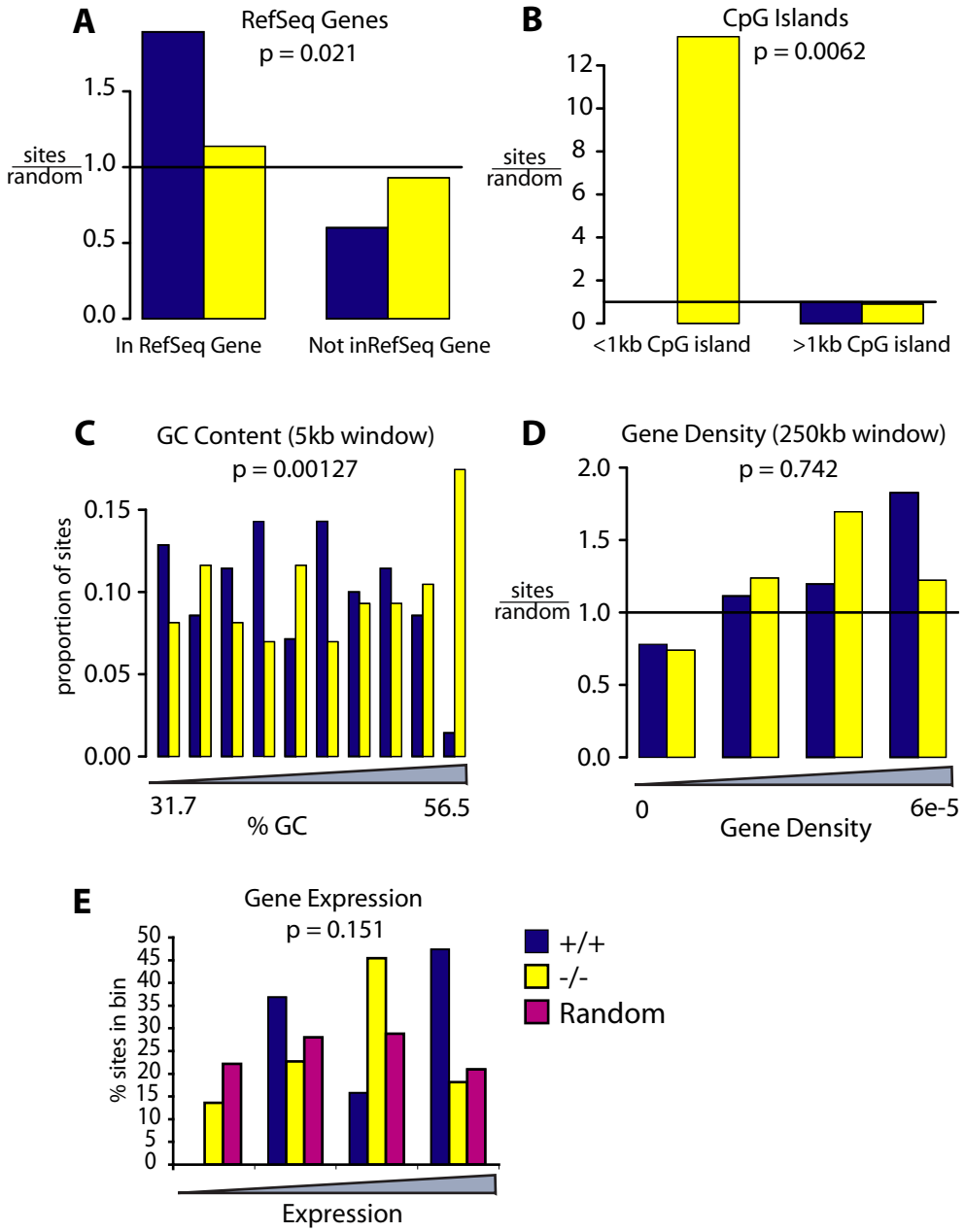

Figure 2-4. EIAV integration distributions in control and LEDGF/p75-disrupted cells. Integration frequencies are shown relative to A) RefSeq genes, B) CpG islands (1 kb window; note that there were no control sites within <1kb), C) G/C content, D) Gene density (250kb window), E) Gene activity. Markings as in Figure 2-3.

As seen above for HIV, the frequency of integration near a variety of features was not detectably altered. Figure 2-4D and E show that the relationships between integration frequency and gene density and expression were not affected by LEDGF/p75 genotype.

Studies of lentiviral integration in human SupT1 cells with intensified knockdown of LEDGF/p75

This work was carried out collaboratively with Heather Marshall, also in the Bushman lab. She carried out studies of HIV and EIAV infection and EIAV integration site selection in a human model of LEDGF/p75 depletion, intensified knockdown SupT1 cells (the TC2 and TL2 cell lines in [37]). Results from human cells were in strong agreement with those from murine cells presented here. HIV and EIAV infection were approximately 10-fold reduced in knockdown cells compared with wild-type or scrambled siRNA-expressing cells (data not shown).

The integration site distribution of EIAV was also altered, and the effects were in agreement with those described above for murine cells. The frequency of integration in transcription units in knockdown cells was significantly reduced compared with control cells, but the proportion of sites in transcription units in knockdown cells remained enriched compared with random (data not shown). Integration within 2kb of CpG islands was also enriched in the human cells (data not shown). In agreement with the data described above, integration preference with respect to gene density and gene expression was not detectably affected by LEDGF/p75 depletion (data not shown).

Correlation between *LEDGF/p75* expression and the frequency of HIV integration in transcription units analyzed over many cell types

In addition to studying cells with artificially reduced levels of *LEDGF/p75* expression, we were interested in natural variation in cellular *LEDGF/p75* expression levels. Different primary cell types and cell lines show different steady state levels of *LEDGF/p75* mRNA. Different cell types also show reproducibly different frequencies of HIV integration in transcription units (see [40] for examples). We thus asked whether cell types with higher LEDGF/p75 levels showed higher frequencies of HIV integration in transcription units.

We analyzed data from 15 HIV integration site data sets for which we also had transcriptional profiling data on gene activity for that cell type. For each microarray data set, the expression level of *LEDGF/p75*-specific probe sets was ranked relative to all other probe sets on the array for that cell type, thus yielding a value for relative *LEDGF/p75* expression. These values were then plotted against the proportion of HIV integration sites in transcription units for that cell type (Figure 2-5). This analysis showed that increased relative *LEDGF/p75* mRNA abundance positively correlated with increased HIV integration frequency in transcription units (R^2=0.61; P<0.0001). Figure 2-5 shows data with experimental LEDGF/p75 knockdowns included (triangles), but the correlation was still significant when the experimental knockdowns were excluded (P<0.0001), indicating that natural variation in LEDGF/p75 levels was functionally significant.

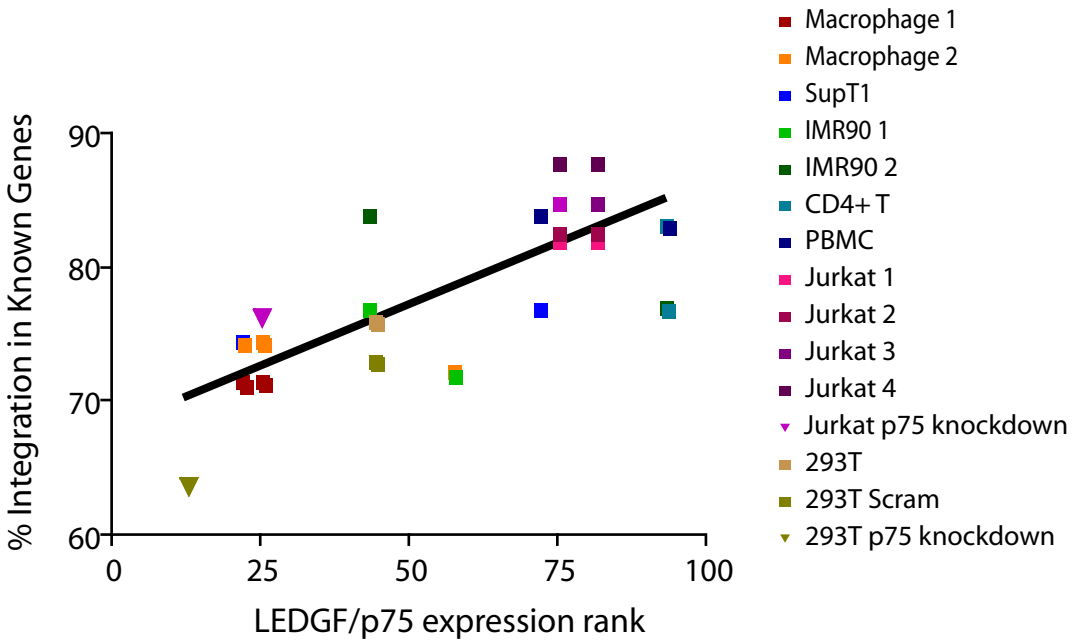

Figure 2-5. Correlation between LEDGF/p75 expression and the frequency of HIV integration in genes. Data is shown for 15 HIV integration site data sets in 10 cell types. The y-axis shows the percentage of integration events within transcription units of the "known gene" set of human genes for each integration site data set. The x-axis shows relative expression values for LEDGF/p75 derived from Affymetrix array data (see methods for details). The R-squared value for the fit is 0.6148 ($P<0.0001$). The references for the data sets used are as follows: Macrophage 1 is the VSV-G set in [25]; Macrophage 2 is the CCR5 set in [25]; SupT1 [21]; IMR90 1 is the dividing set in [66]; IMR90 2 is the growth-arrested set in [66]; CD4 T [67]; PBMC [22]; Jurkat 1 is the Mse set in [46]; Jurkat 2 is the Avr set in [46]; Jurkat 3 is the initially bright set in [5]; Jurkat 4 is the initially dark set in [5]; Jurkat p75 knockdown [40][46]; 293T [40]; 293T Scram [40]; 293T p75 knockdown [40].

Some of the data in Figure 2-5 and in previous studies was generated using transformed cell lines, leaving open the question of whether natural variation in LEDGF/p75 levels was functionally important in human primary cells. We repeated the analysis in Figure 2-5 using only data from human primary cells where LEDGF/p75 levels had not been altered experimentally, and again found a significant positive correlation between integration frequency in genes and *LEDGF/p75* mRNA levels (P=0.044). These data indicate that natural variation in *LEDGF/p75* expression levels is a significant determinant of integration frequency in transcription units in human primary cells.

2.5 Discussion

Here we report studies of lentiviral integration in murine cells with a homozygous gene-trap mutation disrupting the *LEDGF/p75* locus [42]. We present data from HIV and EIAV, extending the collection of lentiviruses shown functionally to be affected by LEDGF/p75. Infectivity for both HIV and EIAV was reduced 5-50 fold in LEDGF/p75-depleted cells, in good agreement with data on HIV and FIV published previously [37, 41] – taken together, these studies firmly establishing that strong LEDGF/p75 knockdowns strongly reduce HIV infectivity. In data not shown, target site selection in human cells closely paralleled the effects in murine cells, and also parallel with studies of another murine *LEDGF/p75* mutant [37, 41].

Published studies of integration targeting by LEDGF/p75 have relied on analysis of cells where the LEDGF/p75 levels were artificially reduced – thus there is interest in obtaining data on the effects of LEDGF/p75 in cells naturally expressing

different levels of the protein. We took advantage of the observation that different cell types differ reproducibly in their frequency of integration in transcription units [40] to investigate this question. A bioinformatic comparison (Figure 2-5) showed that higher levels of *LEDGF/p75* expression correlated with higher frequencies of integration in transcription units. The trend achieved significance even when the analysis was restricted to human primary cells only. Thus the study of natural variation in *LEDGF/p75* expression allowed us to extend the idea that LEDGF/p75 directs HIV integration to transcription units in human primary cells without artificially reduced LEDGF/p75 levels.

A simple model holds that LEDGF/p75 directs favored integration into transcription units by tethering. According to this model, one domain of LEDGF/p75 binds to HIV preintegration complexes and the other binds chromatin at active transcription units. Data from artificial tethering studies *in vitro* with fusions of the LEDGF/p75 IBD to a sequence-specific binding domain support this model [56]. The tethering model predicts that LEDGF/p75 should accumulate on active transcription units, but so far this has not been demonstrated experimentally. Similarly, it is not known how LEDGF/p75 recognizes active transcription units. One possible model would be that histone post-translational modifications mark active transcription units and guide LEDGF/p75 binding. Potentially consistent with this idea is the finding that HIV integration is positively correlated with several types of histone post-translational modifications [46].

Curiously, both this study and Shun *et al.* [41] showed not only a loss of integration targeting in LEDGF/p75-depleted cells, but new favored genomic regions as well. From the previous study alone this might have been an idiosyncrasy, but data presented here shows a similar response in a second murine model, and in human cells. In all LEDGF/p75-depleted cell types in both studies, integration became more favored near transcription start sites and associated CpG islands. The basis for this trend is unknown. It may be that preintegration complexes normally associated with LEDGF/p75 become free to integrate near these sites once LEDGF/p75 was removed. Possibly chromatin at start site regions is particularly accessible and so represents a default target. It is also possible that a more active mechanism is involved. In support of this idea is the finding that MLV integration is strongly favored at start sites [28, 57], while several other integrating elements show near random distributions [22, 55, 58], suggesting that mechanisms exist to guide preferential integration near start sites.

A variety of genomic features showed positive correlation with lentiviral integration in both the depleted cells and controls, indicating that cellular systems in addition to LEDGF/p75 also influence integration. As increasingly deep annotation of the human genome accumulates, it may be possible to detect additional associations between lentiviral integration and particular bound proteins, potentially allowing identification of host cell factors operating in the absence of LEDGF/p75.

Finally, data presented here and in [37, 41] emphasizes that LEDGF/p75 is important for efficient HIV replication, suggesting that the interaction between IN

and LEDGF/p75 may be a tractable target for antiviral therapy. The structure of a complex of the LEDGF/p75 IBD and the IN catalytic domain have been solved by X-ray crystallography [33], and the interaction surface was found to overlap with the binding site seen previously for the integrase inhibitor tetraphenylarsonium [59]. This supports the idea that small molecule inhibitors, if of high enough affinity, may be able to disrupt binding of LEDGF/p75 to integrase and so abrogate HIV replication.

2.6 Acknowledgements

We thank members of the Bickmore, Bushman, and Poeschla laboratory for materials and helpful discussions. We particularly thank Dr. Roderick Jensen and members of the Virginia Bioinformatics Institute Core Laboratory Facility for sequencing support. We also thank John C. Olsen and Manij Patel, at the Cystic Fibrosis/Pulmonary Research and Treatment Center, University of North Carolina at Chapel Hill for the development and donation of the EIAV vectors, maps and infection advice.

2.7 References

1. Coffin JM, Hughes SH, Varmus HE. (1997) Retroviruses. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

2. Bushman FD. (2001) Lateral DNA transfer: Mechanisms and consequences. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

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

4. Jordan A, Bisgrove D, Verdin E. (2003) HIV reporducibly establishes a latent infection after acute infection of T cells ni vitro. EMBO J. 22: 1868-1877.

5. Lewinski M, Bisgrove D, Shinn P, Chen H, Verdin E, et al. (2005) Genome-wide analysis of chromosomal features repressing HIV transcription. J. Virol. 79, 6610-6619.

6. Bisgrove D, Lewinski M, Bushman FD, Verdin E. (2005) Molecular mechanisms of HIV-1 proviral latency. Expert Rev Anti Infect Ther. 3: 805-814.

7. Hacein-Bey-Abina S, von Kalle C, Schmidt M, Le Deist F, Wulffraat N, et al. (2003) A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. N Engl J Med 348: 255-256.

8. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, et al. (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science 302: 415-419.

9. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, et al. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J. Biol. Chem. 278: 372-381.

10. Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, et al. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J. Biol. Chem. 278: 33528-33539.

11. Turlure F, Devroe E, Silver PA, Engelman A. (2004) Human cell proteins and human immunodeficiency virus DNA integration. Front. Biosci. 9: 3187-3208. .

12. Llano M, Vanegas M, Fregoso O, Saenz D, Chung S, et al. (2004) LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. J. Virol. 78: 9524-9537.

13. Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, et al. (2005) Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. J Biol Chem 280: 25517-23.

14. Cherepanov P, Devroe E, Silver PA, Engelman A. (2004) Identification of an evolutionarily-conserved domain of LEDGF/p75 that bind HIV-1 integrase. J. Biol. Chem. 279: 48883-48892.

15. Vanegas M, Llano M, Delgado S, Thompson D, Peretz M, et al. (2005) Identification of the LEDGF/p75 HIV-1 integrase interaction domain and NLS

reveals NLS-independent chromatin tethering. J. Cell. Sci. 118: 1733-1743.

16. Llano M, Vanegas M, Hutchins N, Thompson D, Delgado S, et al. (2006) Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. J Mol Biol 360: 760-773.

17. Turlure F, Maertens G, Rahman S, Cherepanov P, Engelman A. (2006) A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. Nucleic Acids Res 34: 1653-1675.

18. Llano M, Delgado S, Vanegas M, Poeschla EM. (2004) LEDGF/p75 prevents proteasomal degradation of HIV-1 integrase. J. Biol. Chem. 279: 55570-55577.

19. Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, et al. (2005) The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. J Biol Chem 280: 17841-17847.

20. Cherepanov P. (2007) LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. Nucleic Acids Res 35: 113-124.

21. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110: 521-529.

22. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, et al. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol 2: e234.

23. Hematti P, Hong BK, Ferguson C, Adler R, Hanawa H, et al. (2004) Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. PLoS Biol 2: e423.

24. Ciuffi A, Diamond T, Hwang Y, Marshall H, Bushman FD. (2006) Fusions of LEDGF/p75 to lambda repressor promote HIV DNA integration near lambda operators in vitro. Human Gene Therapy 17: 960-967.

25. Barr SD, Ciuffi A, Leipzig J, Shinn P, Ecker JR, et al. (2006) HIV integration site selection: Targeting in macrophages and the effects of different routes of viral entry. Mol Ther 14: 218-225.

26. Barr SD, Leipzig J, Shinn P, Ecker JR, Bushman FD. (2005) Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the

chicken genome. J. Virol. 79: 12035-12044.

27. Crise B, Li Y, Yuan C, Morcock DR, Whitby D, et al. (2005) Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. J Virol 79: 12199-121204.

28. Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, et al. (2006) Retroviral DNA integration: Viral and cellular determinants of target-site selection. PLoS Pathog 2: e60.

29. Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, et al. (2005) Genome-wide analysis of retroviral DNA integration. Nat Rev Microbiol 3:848-858.

30. Kang Y, Moressi CJ, Scheetz TE, Xie L, Tran DT, et al. (2006) Integration site choice of a feline immunodeficiency virus vector. J. Virol. 80: 8820-8823.

31. Hacker CV, Vink CA, Wardell TW, Lee S, Treasure P, et al. (2006) The integration profile of EIAV-based vectors. Mol Ther 14: 536-545.

32. MacNeil A, Sankale JL, Meloni ST, Sarr AD, Mboup S, et al. (2006) Genomic sites of human immunodeficiency virus type 2 (HIV-2) integration: Similarities to HIV-1 in vitro and possible differences in vivo. J Virol 80: 7316-7321.

33. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A. (2005) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc Natl Acad Sci USA 102: 17308-17313.

34. Cherepanov P, Sun ZY, Rahman S, Maertens G, Wagner G, et al. (2005) Solution structure of the HIV integrase-binding domain in LEDGF/p75. Nat. Struct. Mol. Biol. 12: 526-532.

35. Zielske SP, Stevenson M. (2006) Modest but reproducible inhibition of human immunodeficiency virus type 1 infection in macrophages following LEDGFp75 silencing. J Virol 80: 7275-7280.

36. Vandegraaff N, Devroe E, Turlure F, Silver PA, Engelman A. (2006) Biochemical and genetic analyses of integrase-interacting protein lens epithelium-derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication. Virology 346: 415-426.

37. Llano, M., D.T. Saenz, A. Meehan, P. Wongthida, M. Peretz, W.H. Walker, W. Teo, and E.M. Poeschla (2006) An Essential Role for LEDGF/p75 in HIV

Integration. Science 314: 461-464.

38. Vandekerckhove L, Christ F, Van Maele B, De Rijck J, Gijsbers R, et al. (2006) Transient and stable knockdown of the integrase cofactor LEDGF/p75 reveals its role in the replication cycle of human immunodeficiency virus. J Virol 80: 1886-1896.

39. De Rijck J, Vandekerckhove L, Gijsbers R, Hombrouck A, Hendrix J, et al. (2006) Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. J Virol 80: 11498-11509.

40. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, et al. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med 11: 1287-1289.

41. Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, et al. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev 21: 1767-1778. .

42. Sutherland HG, Newton K, Brownstein DG, Holmes MC, Kress C, et al. (2006) Disruption of Ledgf/Psip1 results in perinatal mortality and homeotic skeletal transformations. Mol Cell Biol 26: 7201-7210.

43. Butler SL, Johnson EP, Bushman FD. (2002) HIV cDNA metabolism studied by fluorescence-monitored PCR: Notable stability of two-LTR circles. J. Virol. 76: 3739-3747.

44. Butler S, Hansen M, Bushman FD. (2001) A quantitative assay for HIV cDNA integration in vivo. Nat. Med. 7: 631-634.

45. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376-380.

46. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. (2007) HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res 17: 1186-1194.

47. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW. (1996) Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nat Genet 14: 450-456. .

48. Hoffmann C, Minkah N, Leipzig J, Wang G, Arens MQ, et al. (2007) DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations.

Nucleic Acids Res 35: e91.

49. Binladen J, Gilbert MT, Bollback JP, Panitz F, Bendixen C, et al. (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS ONE 2: e197.

50. Pryciak PM, Sil A, Varmus HE. (1992) Retroviral integration into minichromosomes *in vitro*. EMBO J. 11: 291-303.

51. Stevens SW, Griffith JD. (1996) Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. J. Virol. 70: 6459-6462.

52. Carteau S, Hoffmann C, Bushman FD. (1998) Chromosome structure and HIV-1 cDNA integration: Centromeric alphoid repeats are a disfavored target. J. Virol. 72: 4005-4014.

53. Wu X, Li Y, Crise B, Burgess SM, Munroe DJ. (2005) Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. J. Virol. 79: 5211-5214.

54. Holman AG, Coffin JM. (2005) Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. Proc Natl Acad Sci USA 102: 6103-6107.

55. Berry C, Hannenhalli S, Leipzig J, Bushman FD. (2006) Selection of target sites for mobile DNA integration in the human genome. PLoS Comput Biol 2: e157.

56. Ciuffi A, Diamond TL, Hwang Y, Marshall HM, Bushman FD. (2006) Modulating target site selection during human immunodeficiency virus DNA integration in vitro with an engineered tethering factor. Hum Gene Ther 17: 960-967.

57. Wu X, Li Y, Crise B, Burgess SM. (2003) Transcription start regions in the human genome are favored targets for MLV integration. Science 300: 1749-1751.

58. Miller DG, Trobridge GD, Petek LM, Jacobs MA, Kaul R, et al. (2005) Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. J. Virol. 79: 11434-11442.

59. Molteni V, Greenwald J, Rhodes D, Hwang Y, Kwiatkowski W, et al. (2001) Identification of a small molecule binding site at the dimer interface of the HIV

integrase catalytic domain. Acta Crystallogr. D. Biol. Crystallogr. 57: 536-544.

60. Stone M. (1974) Cross-validatory chosice and assesment of statistical predictions. Journal of the Royal Statistical Society, Series B--Methodological 36: 111-147.

61. Todaro G, Green H. (1963) Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. J Cell Biol 17: 299-313.

62. Follenzi A, Ailes LE, Bakovic S, Gueuna M, Naldini L. (2000) Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. Nat. Genetics 25: 217-222.

63. Naldini L, Blomer U, Gallay P, Ory D, Mulligan R, et al. (1996) In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science 272: 263-267.

64. Olsen JC. (1998) Gene transfer vectors derived from equine infectious anemia virus. Gene Ther 5: 1481-1487.

65. Yamashita M, Emerman M. (2004) Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. J. Virol. 78: 5670-5678.

66. Ciuffi A, Mitchell RS, Hoffmann C, Leipzig J, Shinn P, et al. (2006) Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts. Mol Ther 13: 366-373.

67. Levine BL, Humeau LM, Boyer J, MacGregor RR, Rebello T, et al. (2006) Gene transfer in humans using a conditionally replicating lentiviral vector. Proc Natl Acad Sci USA 103: 17372-17377.

CHAPTER 3 – LEDGF/P75 HYBRIDS RETARGET LENTIVIRAL INTEGRATION INTO HETEROCHROMATIN

The contents of this chapter have been published as part of:

Gijsbers R, Ronen K, Vets S, Malani N, De Rijck J, McNeely M, Bushman FD and Debyser Z. (2010) LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. Mol Ther advance online publication Jan 5, 2010.

3.1 Abstract

Correction of genetic diseases requires integration of the therapeutic gene copy into the genome of patient cells. Retroviruses are commonly used as delivery vehicles because of their precise integration mechanism, but their use has led to adverse events in which vector integration activated proto-oncogenes and contributed to leukemogenesis. Here we show that integration by lentiviral vectors can be targeted away from genes using an artificial tethering factor. During normal lentivirus infection, the host cell encoded transcriptional co-activator LEDGF/p75 binds lentiviral integrase, thereby targeting integration to active transcription units and increasing the efficiency of infection. We replaced the LEDGF/p75 chromatin interaction binding domain with CBX1. CBX1 binds histone H3 di- or tri-methylated on K9, which is associated with pericentric heterochromatin and intergenic regions. The chimeric protein supported efficient transduction of lentiviral vectors and directed integration outside of genes, near bound CBX1. Despite integration in regions rich in epigenetic marks associated with gene silencing, lentiviral vector expression remained efficient. Thus engineered LEDGF/p75 chimeras provide technology for controlling integration site selection by lentiviral vectors.

3.2 Introduction

Lens epithelium-derived growth factor/p75 (LEDGF/p75) is a transcriptional co-activator [1, 2] that colocalizes with chromatin [3] and interacts with the integrase (IN) of the human immunodeficiency virus type 1 (HIV-1) and other *lentivirinae* [4-7]. RNAi-mediated depletion of LEDGF/p75 results in the relocalization of IN to the cytoplasm and blocks HIV replication at the integration step of the viral lifecycle [8-11]. In addition, LEDGF/p75 depletion alters the genomic distribution of lentiviral integration sites [12-14]. Lentiviruses preferentially integrate in active transcription units and disfavor promoter regions and locations within 1kb of CpG islands [13-17]. For both HIV and EIAV (Equine Infectious Anemia Virus), integration in LEDGF/p75-depleted cells is reduced in transcription units, but increased in regions of relatively high GC content and gene 5' ends. A model has therefore been proposed in which LEDGF/p75 functions as a molecular tether, bridging between IN and host chromatin [11-18].

The integrase binding domain (IBD) alone does not mediate chromatin binding, but overexpression relocates HIV IN to the cytoplasm and blocks HIV replication [18,19], likely by blocking function of full length LEDGF/p75. The mechanism of chromatin association by LEDGF/p75 is poorly understood, but an N-terminal PWWP domain [20], a nuclear localization signal, and two AT hooks are implicated as important from functional studies [11, 21-23] (Figure 3-1).

Meehan and colleagues recently showed that LEDGF proteins bearing H1.1, H1.5 and LANA in place of LEDGF's first 199 amino acids are functional HIV-1

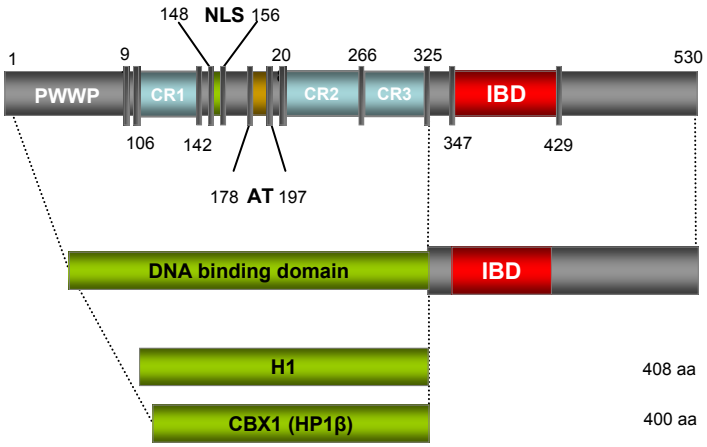

Figure 3-1. Domain structure of LEDGF/p75 and schematic representation of LEDGF$_{325-530}$ fusions. LEDGF/p75 contains an integrase-binding domain (IBD) in the C-terminus and a combination of chromatin interacting modules located in the N-terminal end, most notably the PWWP-domain, the AT-hook domain, and three relatively charged regions (CR1-CR3) influence chromatin binding. In the lower panel the DNA-binding domain fusions with LEDGF$_{325-530}$ are depicted, H1-LEDGF$_{325-530}$ and CBX1-LEDGF$_{325-530}$, respectively. Protein elements are drawn to scale. Numbers indicate amino acids of each domain. NLS, nuclear localization signal; H1, histone H1; CBX1, heterochromatin protein 1β (formerly HP1β).

cofactors [24]. Here we used the LEDGF-IN interaction to retarget lentiviral integration to alternative regions of the genome. We engineered artificial chromatin tethers by fusing the C-terminal IN binding fragment of LEDGF/p75 to alternative chromatin binding proteins, expressed these in LEDGF/p75-depleted cells, and asked whether 1) infection was rescued and 2) integration was retargeted to the regions bound by the chimeric protein. In a previous study, Ciuffi *et al*. created fusions of LEDGF/p75 IBD and the λ repressor DNA binding domain and found increased *in vitro* strand transfer activity near λ repressor binding sites [25]. However, this approach has not yet been used to redirect viral integration in cells.

We compared integration targeting for many hybrids between chromatin binding proteins and LEDGF/p75, with particular focus on domains with binding specificities that might be useful during human gene therapy. The heterochromatin protein 1β (CBX1, formerly HP1β) binds to sites enriched in histone H3K9 di- and tri-methylation at centromeric heterochromatin and transcriptionally silent regions [29, 38, 39]. This provides a chromosomal target present at high copy number in gene sparse regions. We found that a fusion in which CBX1 replaced the chromatin-interaction domain of LEDGF/p75 rescued the infection block in LEDGF/p75-depleted cells. We characterized proviral integration sites using 454 pyrosequencing and found integration to be retargeted in the presence of the fusion to genomic sites bound by CBX1. These regions are preferentially outside transcription units and normally disfavored for lentiviral integration, but transgene expression from the vector was nevertheless efficient. These findings open possibilities for targeting of

gene therapy vectors by using the LEDGF/p75-IN interaction, potentially to gene-poor regions where their genotoxic potential may be reduced.

3.3 Materials and methods

Retroviral vector production and transduction

Lentiviral vector production was performed as described [34, 50]. Briefly, vesicular stomatitis virus glycoprotein (VSV-G) pseudotyped HIV-based particles were produced by PEI transfection using pCHMWS_eGFP-T2A-fLuc as a transfer plasmid [34]. EIAV-vector particles were produced likewise using p6.1G3CeGFPw (M. Patel & J. Olsen, University of North Carolina, Chapel Hill, unpublished) and pEV53B and VSV-G encoding pMD.G.

For lentiviral transduction experiments, cells were typically plated at 20,000 cells/well in a 96-well plate and transduced overnight. 72 hours later, 90% of cells were reseeded into two plates (FACS analysis and Luc-assay). The remainder was cultured for Q-PCR or integration site analysis for at least 20 days to eliminate non-integrated DNA. Stable cell lines were generated by transduction of the monoclonal LEDGF/p75 KD cells with retroviral vectors and subsequent selection with blasticidin (3 µg/ml; Invitrogen, Merelbeke, Belgium).

Q-PCR

Integrated proviral copies were quantified by real-time Q-PCR on gDNA as reported earlier [13].

Integration site amplification

Integration sites were amplified by linker-mediated PCR as described previously [13]. gDNA was digested using MseI and linkers were ligated. Proviral-host junctions were amplified by nested PCR using barcoded primers. This enabled pooling of PCR products into one sequencing reaction. Products were gel-purified and sequenced on the 454 GS-FLX instrument at the University of Pennsylvania. Oligos used are listed below:

Primer	Sequence
EIAV PCR1	CCTGTCTCTAGTTTGTCTGTTCG
EIAV nested PCR2 (454 primer B-barcode-EIAV)	gccttgccagcccgctcagxxxxxxxxAGTTTGTCTGTTCGAGATCCTAC
Linker PCR1	GTAATACGACTCACTATAGGGC
Linker PCR2 (454 primer A-linker)	gcctccctcgcgccatcagAGGGCTCCGCTTAAGGGAC
Mse linker+	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC
Mse linker-	[Phosp]TAGTCCCTTAAGCGGAG-[AmC7-Q]

Bioinformatic analysis

For integration sites to be called as authentic, sequences needed a best unique hit when aligned to the human genome (hg18 draft) using BLAT, the alignment needed to begin within 3bp of the viral LTR end, and the sequence needed to show a >98% match to the human genome. Statistical methods are detailed in Berry *et al* [37]. Integration site counts were compared with matched random controls (MRCs) by a Fisher's exact test (where stated), or by multiple regression models for integration intensity and a c-logit test for significance [37]. Analysis was carried out using R (http://www.r-project.org). Histone modification data from Barski *et al*. [40] and Wang *et al*. [41] were used. The number of sequence tags from the ChIP-Solexa data sets in a defined window around each EIAV integration site or MRC, was calculated. CBX1 binding sites were analyzed using data from Vogel *et al*. [44]. For each DamID probe set available, probes were aligned onto the hg18 draft using

BLAT, and their associated $\log_2$ binding ratios used to select the top 5% of sites. For each integration site or MRC the average number of high-affinity probes within a defined window around the site was calculated. Pericentric regions were defined as 1Mb upstream or downstream of the unsequenced gap on each chromosome.

3.4 Results

Generation of cell lines and LEDGF/p75 fusions

This study was conducted in collaboration with Rik Gijsbers from the Debyser lab. Rik transduced HeLaP4-CCR5 cells with an MLV-based expression vector encoding two miRNA-based shRNAs [26] and a zeocin resistance cassette to generate a monoclonal cell line expressing 4% of parental LEDGF/p75 mRNA.

Rik also generated constructs where LEDGF/p75's chromatin binding region (aa 1-324, Figure 3-1) was replaced by alternative DNA-binding proteins. $LEDGF_{325-530}$ was fused to linker histone 1 (H1; histone 1, H1F0) and heterochromatin protein 1β (CBX1, formerly HP1β). H1F0 binds to nucleosomes without apparent preference for the underlying DNA sequence [27], continuously shuttling among chromatin binding sites [27]. CBX1 is associated with pericentric heterochromatin. CBX1 has a single N-terminal chromodomain which recognizes histone tails via methylated lysine residues, for example tri-methylated histone H3 at lysine 9 (H3K9me3) [29, 38, 39]. Both constructs, referred to as H1-$LEDGF_{325-530}$ and CBX1-$LEDGF_{325-530}$, were introduced in LEDGF/p75 depleted cell lines using MLV-based viral vectors and selected with blasticidin. In parallel, control cell lines complemented with MLV-

based vectors encoding RNAi-resistant LEDGF/p75 (LEDGF BC) or eGFP-$LEDGF_{325-530}$ were generated.

Viability of the selected cell lines was similar to the parental HeLaP4-CCR5 cell line (data not shown). Expression of the fusion proteins in the knockdown cell line (referred to as KD) was verified by Western blot and immunocytochemistry with an antibody against the C-terminal portion of LEDGF/p75 (data not shown). No LEDGF/p75 expression could be detected in KD cells. Back-complementation of KD cells with full-length siRNA-resistant LEDGF/p75 resulted in the expression pattern characterstic of LEDGF/p75, which is dense fine nuclear speckling. Complementation of KD cells with the H1-$LEDGF_{325-530}$ fusion resulted in a nuclear distribution and CBX1-$LEDGF_{325-530}$ which was distributed in multiple irregularly shaped foci over the nuclear area during interphase, a pattern paralleling that of wild-type CBX1 [32,33]. In addition to nuclear localization, LEDGF/p75 fusion proteins were found to mediate chromatin tethering of HIV IN. In accordance with previous data [3,4], transient expression of IN fused to the monomeric red fluorescent protein (mRFP-IN^S) in KD cells resulted in a diffuse fluorescent signal throughout the cytoplasm and complementation with LEDGF/p75 relocated mRFP-IN^S to the nucleus and condensed chromatin. Expression of H1-$LEDGF_{325-530}$ and CBX1-$LEDGF_{325-530}$ rescued the nuclear localization of mRFP-IN^S and the binding to condensed chromatin (data not shown).

LEDGF hybrids rescue lentiviral transduction

After demonstrating that the fusions were capable of interacting with HIV-1 IN and tethering IN to chromatin, Rik Gijsbers also assayed the efficiency of integration by lentiviral vectors. In addition to interacting with HIV IN, LEDGF/p75 is known to interact with other lentiviral integrases [4-5]. Integrated HIV- and EIAV-based vector proviral copies were quantified in the different cell lines. For HIV, KD cells showed a 5.8-fold decrease in integrated copies compared with wild-type cells, which was rescued completely upon back-complementation with full-length RNAi-resistant LEDGF/p75. Expression of fusion proteins partially rescued integration (60% and 41% of LEDGF BC integration for H1- and CBX1-LEDGF$_{325-530}$, respectively). Similarly, for EIAV, the number of integrated copies in KD cells was decreased 8.8-fold compared to wild-type cells. Complementing the KD cells with H1-LEDGF$_{325-530}$ and CBX1-LEDGF$_{325-530}$ resulted in a partial rescue of vector integration (3.3-fold and 6.9-fold increase over KD respectively). Thus expression of the chimeric proteins partially rescued EIAV and HIV integration.

Sequencing of proviral integration sites

We next asked whether the LEDGF$_{325-530}$ fusions retargeted integration to genomic sites bound by the fusion partner. Since HeLaP4 cells contain integrated HIV LTRs that would interfere with the isolation of HIV provirus, we used the EIAV vector for distribution analysis. EIAV and HIV integrase both interact with the LEDGF/p75 IBD [6] and show the same integration site preferences in wild-type [35] or LEDGF/p75-depleted [13] cells. Integration sites were analyzed as described previously [13], yielding a total of 2769 integration sites (Table 3-1). Random control

	Cell line	Number of sites	% in RefSeq genes	% <2kb CpG Island
EIAV sites	Wild-type	717	67.2***	1.3
	Back-complemented	862	70.2***	1.9
	Knockdown	213	51.2***	5.6**
	H1-LEDGF$_{325-530}$	449	46.1**	3.3
	CBX1-LEDGF$_{325-530}$	528	32.6*	1.1
MRC sites	MRC WT	2151	37.3	2.8
	MRC BC	2586	36.9	2.1
	MRC KD	639	36.5	1.9
	MRC H1	1347	36.8	2.2
	MRC CBX1	1584	37.8	2.1

Table 3-1. Integration sets generated in this study and their genomic distributions. Significant deviation from matched random controls (MRC) according to the Fisher's exact test is denoted (*** $p<0.0001$, ** $p<0.01$, * $p<0.05$).

sites were generated computationally, and matched to experimental sites with respect to the distance to the nearest MseI cleavage site (matched random control, MRC). In the analyses that follow, the distribution of experimental EIAV sites is normalized to that of the MRC sites, as a control for recovery bias due to cleavage by restriction enzymes [36, 37].

Retroviral integrases show weak but detectable target sequence specificity at the local site of integration. In line with previous reports [13,14], LEDGF/p75 depletion did not affect the consensus sequence flanking the integration site (Figure 3-2). Likewise, expression of $LEDGF_{325-530}$ fusions did not alter the consensus sequence, consistent with the idea that IN binding to local target DNA determines the sequence preference, independently of the tethering mechanism.

CBX1-fusion directs integration to intergenic regions

Lentiviruses favor integration in transcription units and gene-dense regions [15, 35]. In the absence of LEDGF/p75 this preference is reduced, and a preference for CpG islands and gene 5' ends emerges [12-14]. As an initial survey of the proviral integration site distribution, we examined the frequency of integration in these features. In KD cells a reduction in the integration frequency in RefSeq transcription units from 67.2% to 51.2% was observed (Table 3-1 and Figure 3-3A), as previously reported for LEDGF/p75-depleted cells [13]. While this reduction was statistically significant ($p<0.0001$, Fisher's exact test), integration events in the KD cells were still significantly favored in transcription units over random ($p=4.8$ e-6). In accordance with previous reports, we found that integration sites in the KD cells were

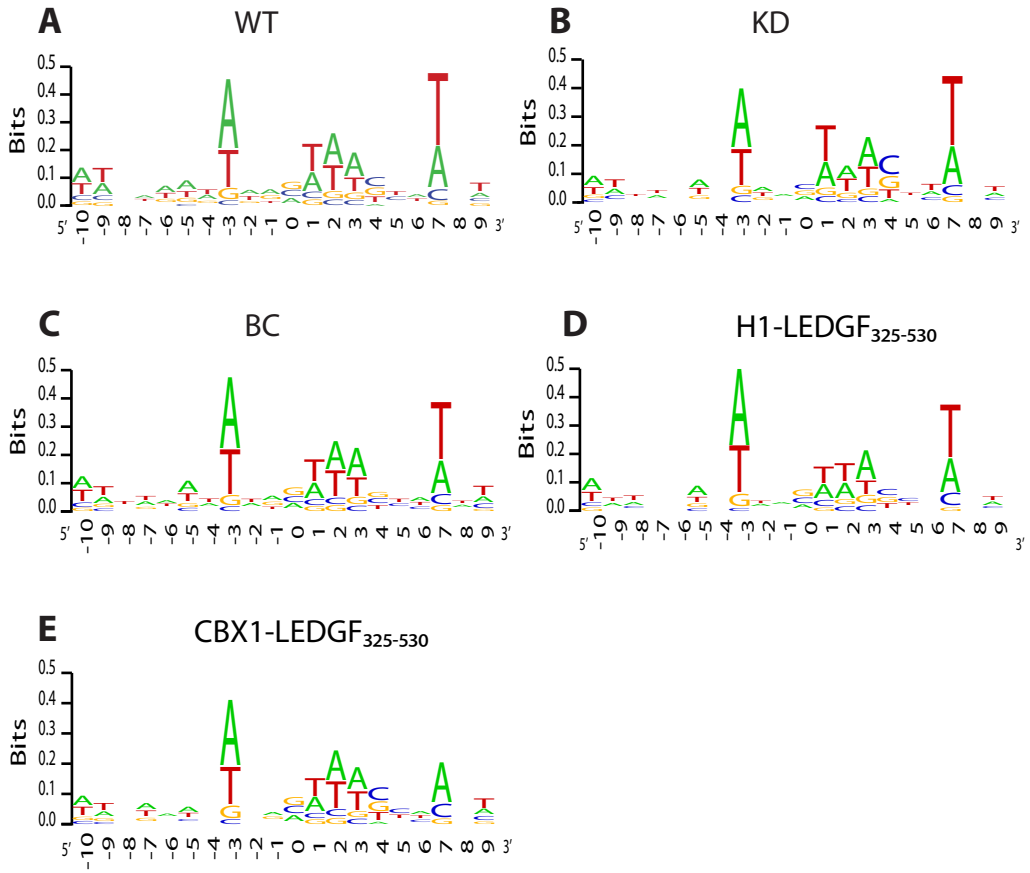

Figure 3-2. EIAV integration site consensus is not affected by LEDGF/p75 fusions. Following alignment of 20 bp surrounding EIAV integration sites from each cell type, a consensus sequence was generated using the WebLogo program (weblogo.berkeley.edu/logo/cgi). Consensus sequence for integration sites from **A.** WT cells **B.** KD cells **C.** cells complemented with full-length LEDGF/p75 **D.** cells stably expressing H1-LEDGF325-530 **E.** cells stably expressing CBX1-LEDGF325-530. The x-axis shows the position relative to the integration site (between position -1 and 0). The y-axis shows bits of information at each position (perfect conservation at a position would score 2 bits).

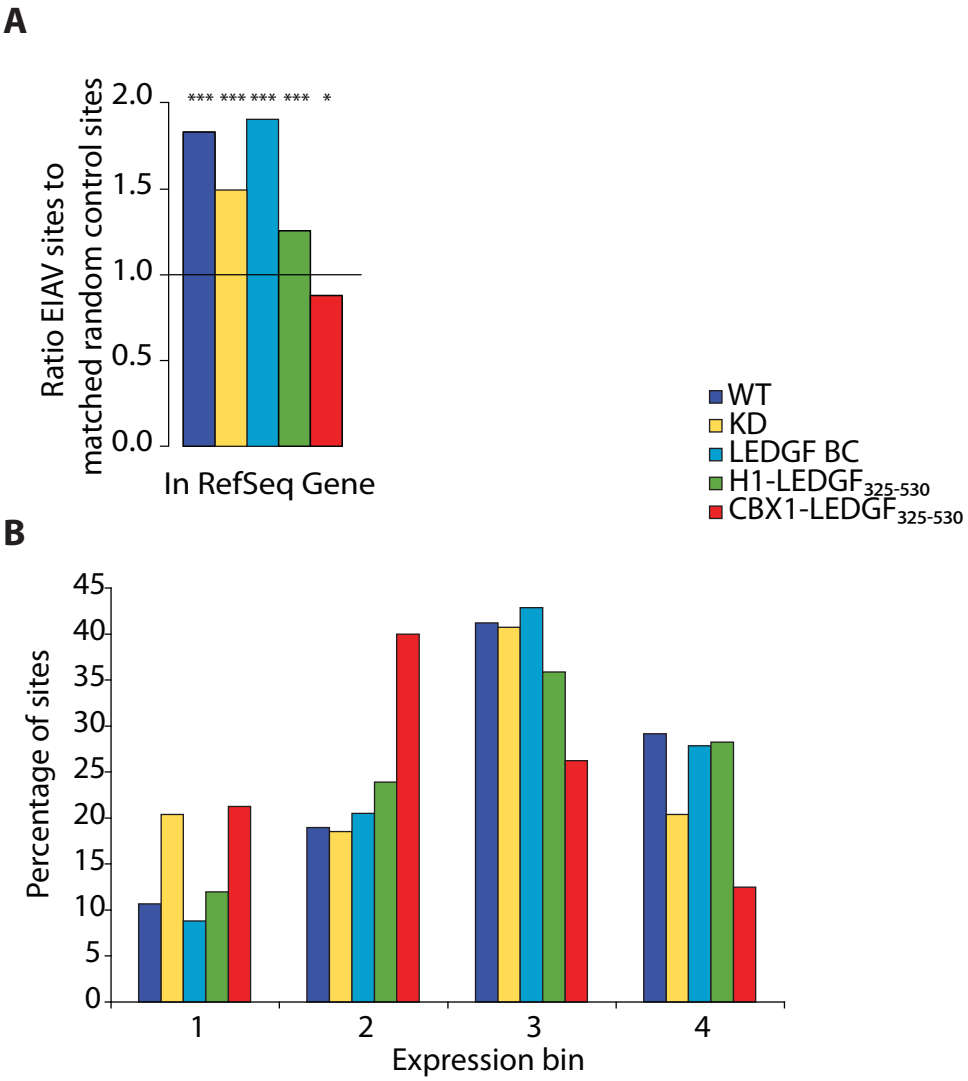

Figure 3-3. Expression of the CBX1 fusion retargets EIAV integration away from expressed genes. A. Frequency of integration within transcription units, as defined by the RefSeq gene call. Values are normalized to matched random controls. Asterisks signify significant deviation from matched random controls, determined by the Fisher's exact test. *p<0.05, **p<0.01, ***p<0.0001 **B.** Distribution of integration sites with respect to gene expression. Transcriptional profiling data from wild-type HeLa cells analyzed on the HU133 Affymetrix microarray were used to classify the expression of genes harboring EIAV integration sites. Genes represented on the microarray were ranked by expression level and divided into 4 bins. For each cell line, integration sites found within genes were sorted into the 4 bins. The percentages shown are the proprtions of sites in each bin out of sites falling in all expression bins.

favored near CpG islands. Both trends were reversed by LEDGF/p75 back-complementation. In contrast, expression of H1-LEDGF$_{325-530}$ did not rescue integration in transcription units. However, upon expression of CBX1-LEDGF$_{325-530}$, integration was significantly disfavored in transcription units compared with random ($p=0.026$, Fisher's exact test), consistent with the distribution pattern of CBX1 in heterochromatic regions, which are generally gene-poor. Analysis of the distribution of integration sites with respect to gene expression level of the targeted genes also provided evidence of retargeting due to expression of the CBX1-LEDGF fusion. Figure 3-3B shows that EIAV integration sites falling in transcription units showed a slight shift towards genes with lower expression level according to microarray analysis of HeLa cells ($p<0.0001$, comparing CBX1-LEDGF$_{325-530}$ and WT cells by the Chi Square test for trend).

CBX1-fusion directs integration to heterochromatic regions

CBX1 is known to bind H3 di- or tri-methylated at K9 (H3K9me2 and H3K9me3, respectively) via its chromodomain [29, 38, 39], so we investigated integration near sites of these histone modifications [40, 41]. The H3K9me3 density near sites of EIAV integration is summarized in Figure 3-4A. In WT cells, integration was disfavored in areas high in H3K9me3 ($p=2.9$ e-29), consistent with the role of H3K9me3 in transcriptional repression and establishment of silent heterochromatin, features generally disfavored by lentiviral integration. In the KD cells the same negative correlation remained, though its magnitude was reduced ($p=0.0012$). Complementation with LEDGF/p75 restored the negative effect of H3K9me3 to wild-

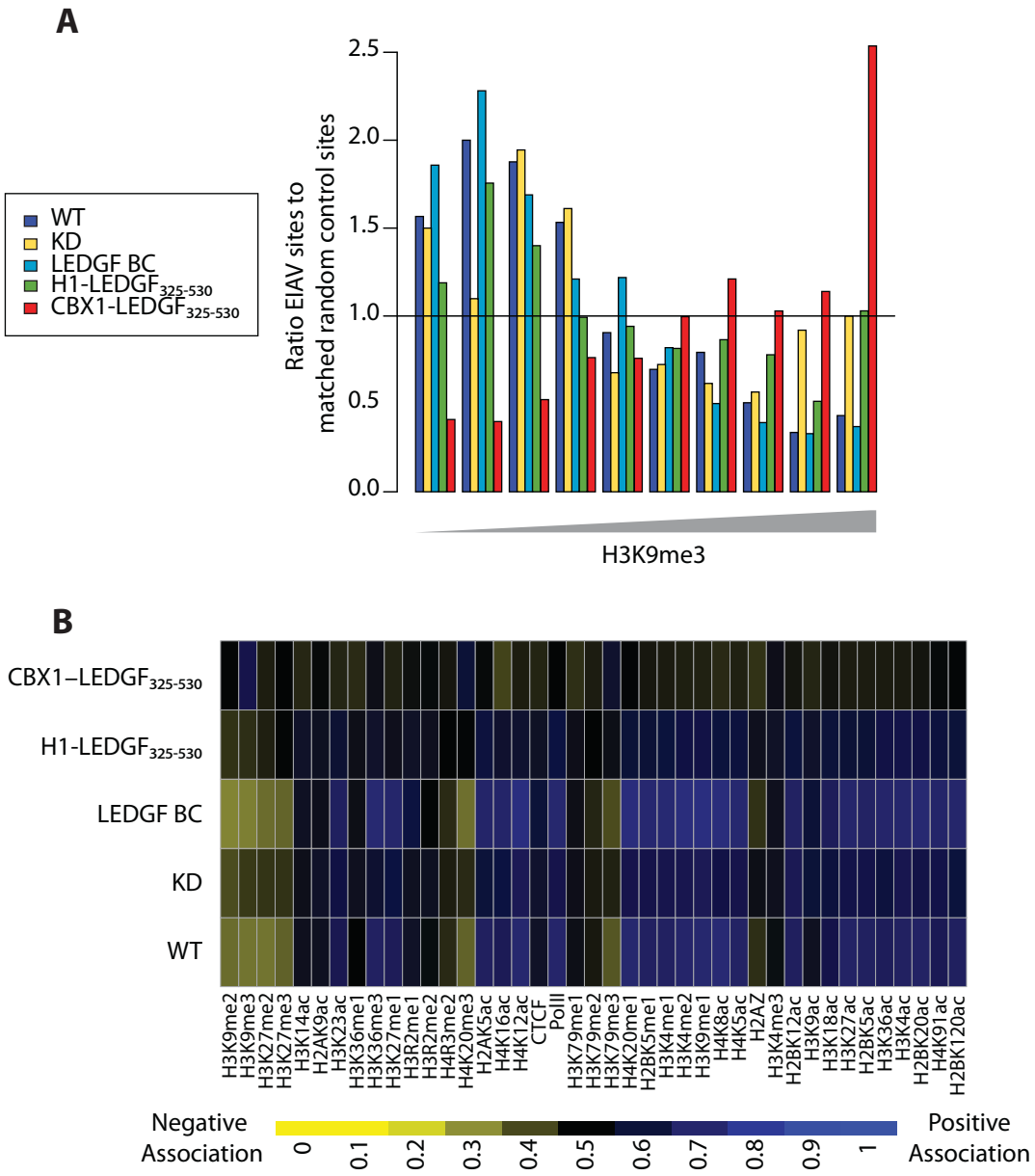

Figure 3-4. Expression of the CBX1 fusion retargets EIAV integration into CBX1-rich heterochromatin regions. A. Relationship of integration frequency to sites of H3K9me3 (10kb window around each site). **B.** Integration frequency relative to density of histone methylation and acetylation density (10kb window around sites). **C.** Integration frequency in pericentromeric regions (defined as 1 Mb at the edge of unsequenced centromeric regions). **D.** Integration frequency in human chromosome 19 near CBX1 binding sites. See Materials and Methods for detailed explanation of anaylsis.

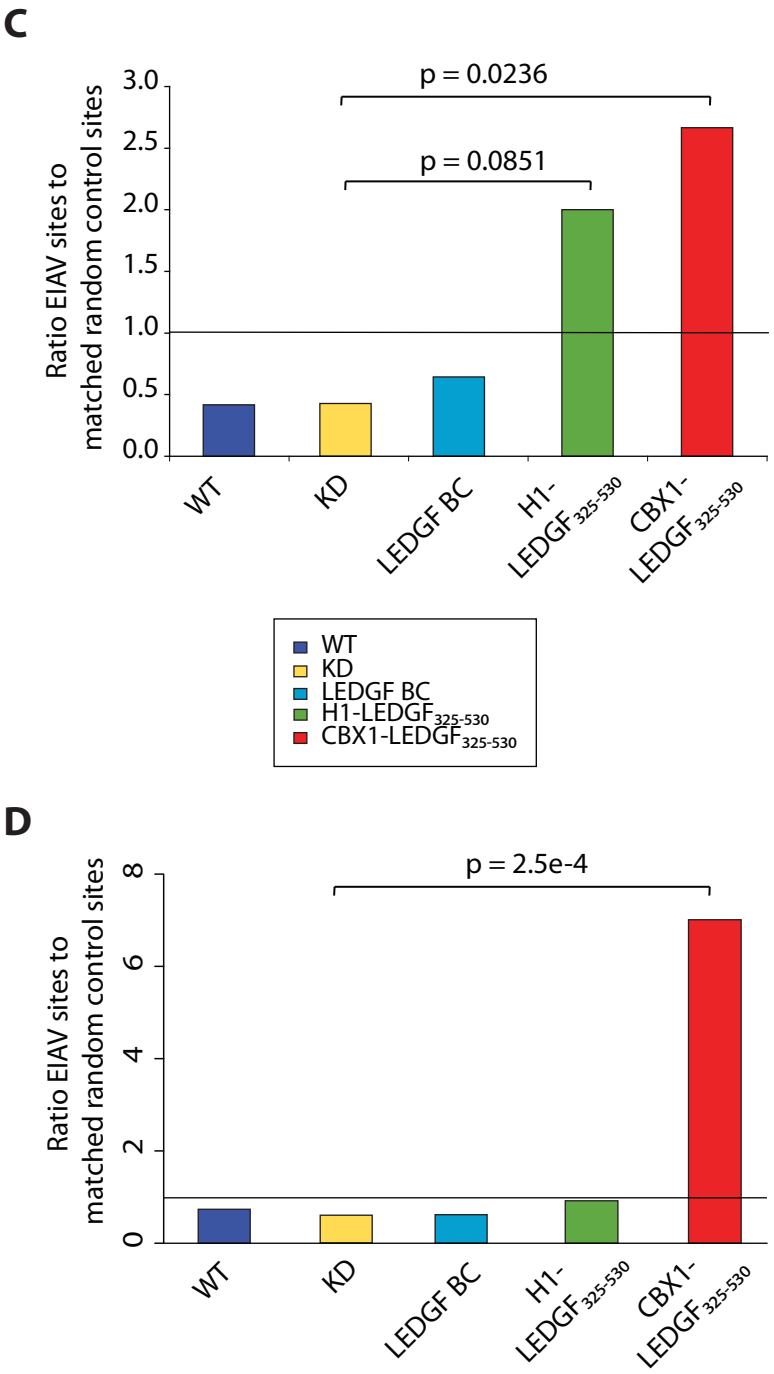

Figure 3-4. continued

type levels. Integration site distribution in H1-LEDGF$_{325-530}$ cells paralleled that seen in KD cells. In cells expressing CBX1-LEDGF$_{325-530}$, however, the correlation was reversed, with integration sites showing a clear preference for regions denser in H3K9me3 (p=1.3 e-13).

We carried out the same analysis using genome-wide ChIP-seq data for a panel of 39 histone modifications [41]. Figure 3-4B shows correlations between integration sites and the density of these modifications. Each correlation is represented as a tile on the heat map, with the color denoting the strength and direction of the correlation. Histone modifications are grouped into clusters, reported to colocalize and associate with classes of functional genomic elements [41]. In wild-type cells, EIAV sites positively correlated with histone modifications generally associated with active transcription, such as all acetylations, and some histone methylations (shown in blue). Integration sites in wild-type cells negatively correlated (shown in yellow) with H3K9me3 and other markers reported to be associated with transcriptionally silent regions (e.g. H3K27me3) and heterochromatin (e.g. H4K20me3 and H3K79me3) [40, 42, 43]. In KD cells, most of the correlations persisted, though they were less pronounced. Complementation with LEDGF restored correlations to wild-type levels. In cells expressing CBX1-LEDGF$_{325-530}$, however, most of the correlations were reversed, suggesting a dramatic redistribution of integration sites. In addition to H3K9me3, the modification bound by CBX1, regions high in H4K20me3 and H3K79me3 became favored for EIAV integration. The latter two modifications have also been associated with pericentric heterochromatin.

Since CBX1 is enriched around centromeres, we compared the frequency of integration sites in pericentric regions. Integration sites in WT, KD or LEDGF BC cells did not differ from random (Figure 3-4C). In contrast, in cells expressing CBX1-$LEDGF_{325-530}$, these regions contained 2.7-fold as many integration sites as MRC sites (p=0.0052, Fisher's exact test), significantly higher than KD cells (p=0.0236, Fisher's exact test). Sites from H1-$LEDGF_{325-530}$ cells also showed a preference for these regions, but this was not significantly higher than KD cells (p=0.0851, Fisher's exact test).

Finally, we used CBX1 binding sites mapped by DamID [44] to calculate the average number of CBX1 binding sites around integration sites. Figure 3-4D shows CBX1 occupancy around EIAV integration sites on chromosome 19 did not differ from random in wild-type cells and KD cells, and was not altered in H1-$LEDGF_{325-530}$ expressing cells. However, in cells complemented with the CBX1 fusion, 10kb windows around integration sites contain 7 times as many CBX1 binding sites as random (p=2.5 e-4). The same pattern held when integration sites across the genome were compared to CBX1 binding sites mapped genome-wide (p=0.015, not shown). Thus the CBX1-LEDGF/p75 fusion redirected integration to sites known to bind CBX1 and a collection of associated features.

Reporter gene expression remains efficient over time

Having shown that the CBX1-$LEDGF_{325-530}$ fusion retargets lentiviral integration to sites bound by CBX1, we wondered whether gene expression from the vector remains efficient, despite integration in regions rich in epigenetic marks

associated with gene silencing. Heterochromatin is known to spread to surrounding chromatin – indeed CBX1 and the other HP1 isoforms are thought to enable this propagation, by binding trimethylated H3K9 and recruiting the methyltransferase responsible for depositing the modification, Suv39 [29]. This is thought to account for the phenomenon of position-effect variegation, an effect first described in *Drosophila*, where chromosome rearrangements resulting in the positioning of a gene adjacent to centromeric heterochromatin resulted in its silencing (reviewed in [51]). We therefore wondered whether targeting lentiviral integrants to regions bound by the heterochromatin-associated protein CBX1 would have a similar silencing effect.

Cells lines were infected with an HIV vector expressing luciferase and luciferase activity measured by Rik Gijsbers 48h post-infection. Luciferase activity was 7-fold lower in KD cells than WT (data not shown) and was rescued by back-complementation with RNAi-resistant LEDGF. Fusion of $LEDGF_{325-530}$ to either the linker histone H1 or CBX1 partially rescued viral vector transduction (36.3% and 47.5%, respectively, data not shown). Similarly for EIAV, back-complementation of KD cells rescued vector transgene expression to wild-type levels and fusion of $LEDGF_{325-530}$ to linker histone H1 or CBX1 partially rescued viral expression (53% and 45.1%, respectively, data not shown). Thus, partial reporter gene expression was observed, mirroring the partial integration rescue observed by provirus QPCR.

We wondered whether the integrants in cells expressing CBX1-$LEDGF_{325-530}$ might be silenced over time due to the spread of inhibitory chromatin. Reporter activity was therefore measured over time for the CBX1-$LEDGF_{325-530}$ cells and

compared to WT cells, KD cells or KD cells complemented with full-length LEDGF/p75. Engineered cell lines were infected with an HIV-based vector expressing eGFP and fLuc [34], and reporter expression was measured in cells over two weeks. Mean fluorescence intensity (MFI) gradually decreased in all cell lines over time (Figure 3-5). The relative difference in overall MFI (fold difference to the first measurement at 72 hrs) reached 4-fold in WT cells or LEDGF BC cells, and about 2.5-fold in the KD cells. Surprisingly, eGFP reporter activity only decreased 1.5-fold in the CBX1-LEDGF$_{325-530}$ cells, demonstrating that despite retargeting to CBX1 binding regions transgene expression remained efficient. Thus, there was no evidence for the idea that integration in more heterochromatic regions directed by the CBX1 domain obstructed gene expression from these HIV-based vectors.

In these studies of lentiviral vectors, reporter gene expression is driven by the CMV immediate early promoter. We wondered if a similar effect would be observed with LTR promoter elements. Cells were therefore infected with replication-competent HIV (NL4.3 strain) and p24 production quantified over time by ELISA (Figure 3-6). Again, CBX1-LEDGF$_{325-530}$-expressing cells mediated a partial rescue of Gag gene expression, to 26.2% of that in WT cells. However, in QPCR-based assays CBX1-LEDGF$_{325-530}$ expression mediated 60% rescue of integration, suggesting that in this case vector expression efficiency may be reduced. It therefore remains possible that retargeting lentiviral integration to heterochromatic regions does influence vector expression, though the effects appear to be modest and promoter-specific.

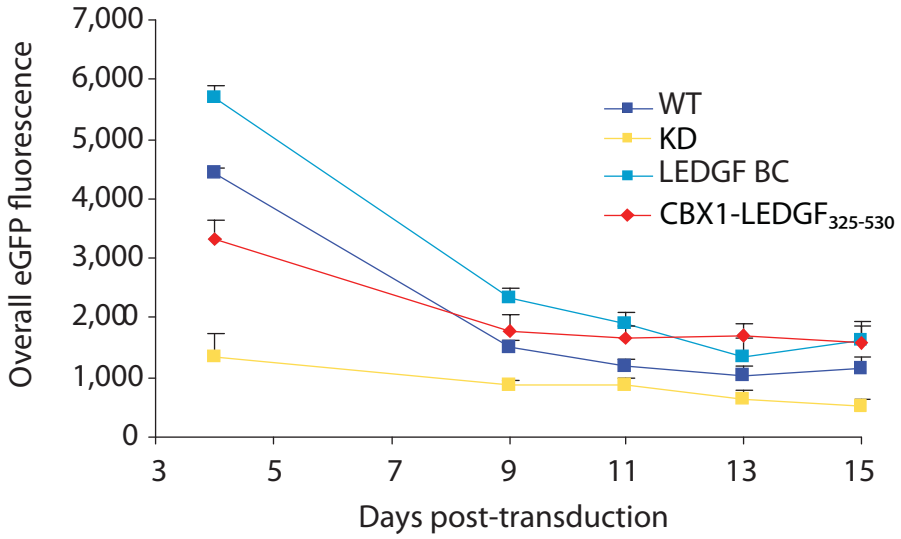

Figure 3-5. Effect of retargeting by CBX1-LEDGF$_{325-530}$ on transgene expression over time. WT, KD and LEDGF BC cells were used as controls. GFP expression was assayed at the indicated time point following HIV-based vector (LV CMV eGFP-T2A-fLuc) infection (dpi, days post infection). Overall eGFP fluorescence over time is calculated as MFI multiplied by % gated cells and displayed as mean ± stdev (n=6). Experiment conducted by Rik Gijsbers.

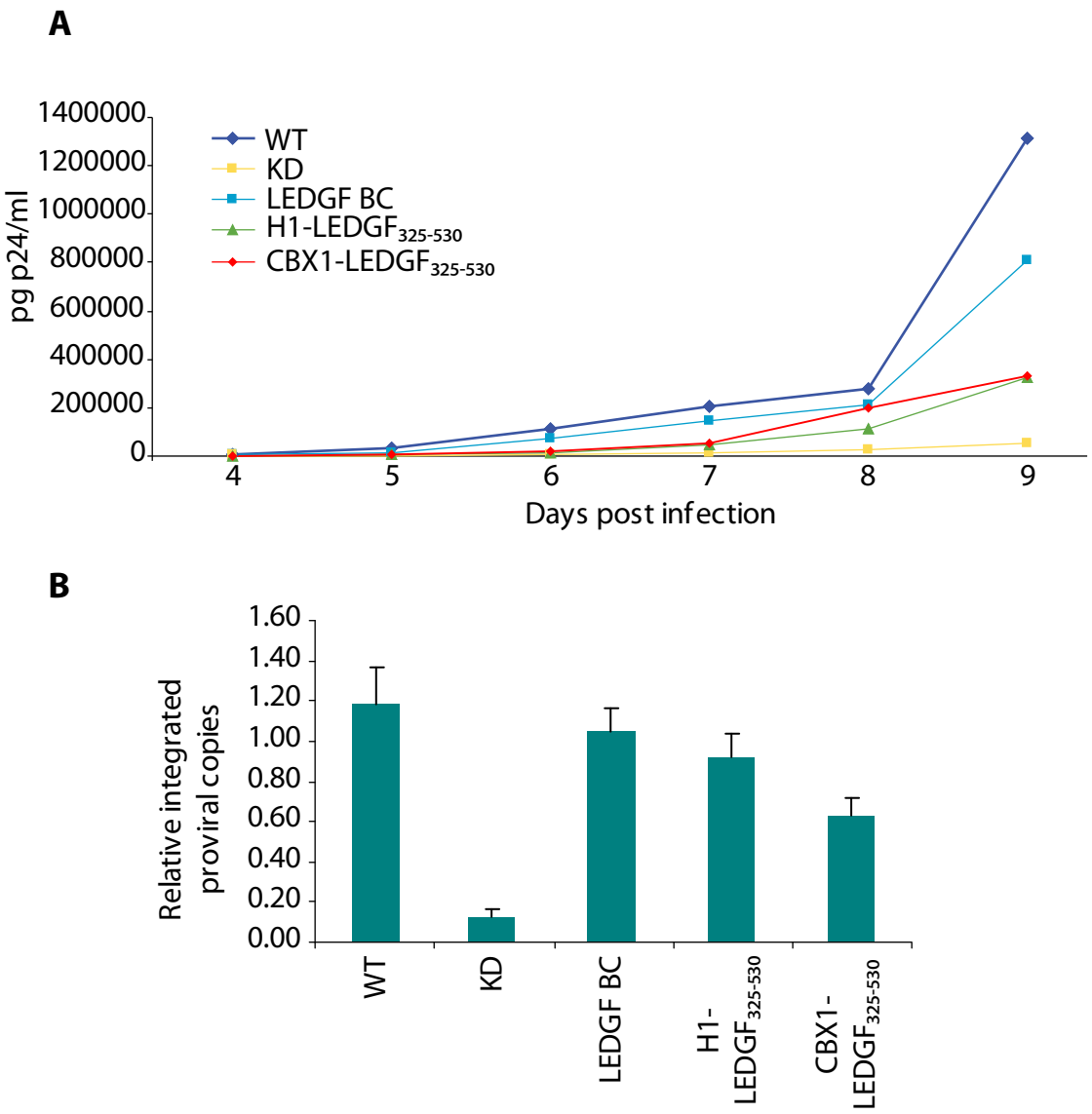

Figure 3-6. Effect of integration retargeting on HIV (NL4.3) gene expression. **A.** Gag p24 expression, quantified by ELISA at intervals post-infection. **B.** Proviral copies in cells expressing the different constructs, normalized to the level in back-complemented cells. Experiments conducted by Rik Gijsbers.

3.5 Discussion

In this study, we present evidence that LEDGF/p75 can be engineered to target lentiviral integration to new positions in the genome. Alternative chromatin binding domains (linker histone H1 or the heterochromatin protein 1β, CBX1, were fused to the C-terminal portion of LEDGF/p75 (aa 325-530, $LEDGF_{325-530}$). CBX1 was selected to target sites of H3K9 di- and tri-methylation, which are mapped in the genome and usually disfavored for lentiviral integration, so retargeting would be readily identifiable. H1 was used as a control, since it has no known preference for the underlying DNA sequence. Fusing a new chromatin-binding module to $LEDGF_{325-530}$ changed the behavior of this protein from an integration-inhibitor into an efficient cofactor. Upon challenge by lentiviral vectors, $LEDGF_{325-530}$-fusions supported efficient lentiviral transduction and integration compared to KD cells. Similar data were recently reported by Meehan and co-workers [24], albeit using LEDGF-hybrids that only lack the PWWP- and AT-hook domain (aa 1-199).

In addition, we characterised proviral integration sites using 454 pyrosequencing. Analysis of the EIAV integration distribution demonstrated that the CBX1 fusion retargeted lentiviral integration away from RefSeq genes (Table 3-1), to regions high in H3K9me3 (Figure 3-4A) and CBX1 binding (Figure 3-4D). The observation that integration can be retargeted away from genes and into heterochromatin using LEDGF hybrids raises hope for the development of safer lentiviral vectors for gene therapy. Prior to this study, attempts to retarget HIV integration employed fusions of IN with DNA-binding proteins [45-48]. Some of

these showed retargeting as purified enzymes, but until now this approach had limited effect on the distribution of integration sites in cells.

The CBX1 hybrid provides the first example of global redistribution of lentiviral integration sites in the cellular genome, and the first instance of manipulation of a host tethering factor to do so. The success of the CBX1 fusion may be due to the abundance in the genome of its target ligand compared with site-specific DNA binding domains previously employed, or perhaps its level of occupancy.

Even though integration was targeted towards regions in the genome that are generally associated with gene silencing, transgene expression remained efficient over time (Figure 3-5). Similarly, when the effect of retargeting on a replication-competent HIV strain, NL4.3, was tested (where gene expression is driven by LTR promoter elements) Gag-p24 gene expression and infection spread still occurred, though there was some evidence that its efficiency might be reduced. This suggests retargeting integration to heterochromatin may have promoter-specific effects. Based on the paradigm of position effect variegation, one might have expected an effect on reporter gene expression [51]. The idea that the genomic location of an integrated HIV provirus affects its expression would also be consistent with previous studies and proposed models of HIV latency [52]. These results therefore warrant further investigation. It may be that expression of the CBX1-LEDGF/p75 fusion protein interferes with the spread of heterochromatin mediated by endogenous CBX1. Whether new classes of genes are activated as a result is unknown. Alternatively it may be that lentiviral LTRs contain unidentified insulator elements. To determine the

chromatin structure of integrated proviruses, chromatin immunoprecipitation of regions of the provirus could be performed. Another way to determine if proviral expression is reduced by retargeting would be to employ the approach of Lewinski *et al.* [52], who looked at genomic features correlating with inducible (instead of constitutive) provirus expression. Cells were infected with GFP Tat-dependent virus, selected for stably bright cells and dim but inducible cells and integration site distributions were compared. This could be used to ask if proviruses that express poorly are also viruses that show more extreme retargeting.

Our findings open possibilities to engineer viral vectors that incorporate LEDGF/p75 hybrids to target integration into safe landing sites, thereby reducing the risk of insertional mutagenesis. Hare and colleagues have recently reported [49] a set of amino-acid substitutions in HIV IN that abolish LEDGF/p75 binding, together with mutations in the LEDGF/p75 protein that restore binding. Gene delivery vectors could thus use an altered IN/LEDGF pair to direct integration, even in the presence of wild-type LEDGF/p75. To date the altered IN does not show wild-type integration activity, but this may be improved with further engineering.

Our data also address issues in HIV biology. Our findings strengthen the idea that LEDGF/p75 is the dominant tether for lentiviral integration. The fact that integration can be retargeted to genomic regions usually disfavored for integration indicates that integration in these areas in wild-type cells is disfavored due to the lack of a tether, rather than to an inherent integration barrier such as steric hindrance resulting from the condensed chromatin structure. Moreover, we show that

chromatin-binding proteins with multiple specificities can successfully replace the LEDGF/p75 DNA-binding elements and rescue HIV infection in a LEDGF/p75 knockdown model. Still, the hybrids did not mediate rescue to wild-type levels, which leaves open the question of whether some portions of the N-terminus of LEDGF/p75 absent from our fusions stimulate integrase activity or reporter gene expression. Indeed, a recent study suggested that serine residues 271, 273 and 275 may be important in LEDGF/p75 cofactor function without affecting DNA or integrase binding [53].

In conclusion, these results establish that LEDGF/p75 is the dominant targeting factor for lentiviral integration and that its interaction with lentiviral integrases can be exploited to develop safe and target-specific lentiviral vectors for gene therapy.

3.6 Acknowledgements

We thank Rik Gisjbers and members of the Debyser lab for their collaboration. Rik Gijsbers is a postdoctoral fellow of the Flemish Fund for Scientific Research (FWO Vlaanderen). Keshet Ronen was supported by NIH training grant T32 AI-07324-17. Sofie Vets is funded by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). Jan De Rijck has a Mathilde-Krim postdoctoral fellowship from Amfar. Melissa McNeely is funded by the Affiliation Française contre les Maladies Musculaire (AFM). Research was funded by the CellCoVir SBO grant (60813) of the IWT; FWO grant G.0530.08; EC grant THINC (HEALTH-F3-2008-201032) to Zeger Debyser and by NIH grant

AI52845, the University of Pennsylvania Center for AIDS Research, and the Penn Genome Frontiers Institute with a grant with the Pennsylvania Department of Health to Frederic D. Bushman.

3.7 References

1. Dietz F, Franken S, Yoshida K, Nakamura H, Kappler J, Gieselmann V. The family of hepatoma-derived growth factor proteins: characterization of a new member HRP-4 and classification of its subfamilies. *The Biochemical journal* 2002; **366**(Pt 2)**:** 491-500.

2. Shinohara T, Singh DP, Fatma N. LEDGF, a survival factor, activates stress-related genes. *Progress in retinal and eye research* 2002; **21**(3)**:** 341-358.

3. Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, Debyser Z *et al.* LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *The Journal of biological chemistry* 2003; **278**(35)**:** 33528-33539.

4. Llano M, Vanegas M, Fregoso O, Saenz D, Chung S, Peretz M *et al.* LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *Journal of virology* 2004; **78**(17)**:** 9524-9537.

5. Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, Christ F *et al.* The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *The Journal of biological chemistry* 2005; **280**(18)**:** 17841-17847.

6. Cherepanov P. LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic acids research* 2007; **35**(1)**:** 113-124.

7. Cherepanov P, Devroe E, Silver PA, Engelman A. Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *The Journal of biological chemistry* 2004; **279**(47)**:** 48883-48892.

8. Maertens G, Cherepanov P, Debyser Z, Engelborghs Y, Engelman A. Identification and characterization of a functional nuclear localization signal in the HIV-1 integrase interactor LEDGF/p75. *The Journal of biological chemistry* 2004; **279**(32)**:** 33421-33429.

9. Vanegas M, Llano M, Delgado S, Thompson D, Peretz M, Poeschla E. Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering. *Journal of cell science* 2005; **118**(Pt 8)**:** 1733-1743.

10. Vandekerckhove L, Christ F, Van Maele B, De Rijck J, Gijsbers R, Van den Haute C *et al.* Transient and stable knockdown of the integrase cofactor LEDGF/p75 reveals its role in the replication cycle of human immunodeficiency virus. *Journal of virology* 2006; **80**(4)**:** 1886-1896.

11. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH *et al.* An essential role for LEDGF/p75 in HIV integration. *Science (New York, NY* 2006; **314**(5798)**:** 461-464.

12. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P *et al.* A role for LEDGF/p75 in targeting HIV DNA integration. *Nature medicine* 2005; **11**(12)**:** 1287-1289.

13. Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, Saenz D *et al.* Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS ONE* 2007; **2**(12)**:** e1340.

14. Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, Kellam P *et al.* LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes & development* 2007; **21**(14)**:** 1767-1778.

15. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 2002; **110**(4)**:** 521-529.

16. Ciuffi A, Bushman FD. Retroviral DNA integration: HIV and the role of LEDGF/p75. *Trends Genet* 2006; **22**(7)**:** 388-395.

17. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome research* 2007; **17**(8)**:** 1186-1194.

18. Hombrouck A, De Rijck J, Hendrix J, Vandekerckhove L, Voet A, De Maeyer M *et al.* Virus evolution reveals an exclusive role for LEDGF/p75 in chromosomal tethering of HIV. *PLoS pathogens* 2007; **3**(3)**:** e47.

19. De Rijck J, Vandekerckhove L, Gijsbers R, Hombrouck A, Hendrix J, Vercammen J *et al.* Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. *Journal of virology* 2006; **80**(23)**:** 11498-11509.

20. Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends in biochemical sciences* 2003; **28**(2)**:** 69-74.

21. Turlure F, Maertens G, Rahman S, Cherepanov P, Engelman A. A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. *Nucleic acids research* 2006; **34**(5)**:** 1653-1675.

22. Llano M, Vanegas M, Hutchins N, Thompson D, Delgado S, Poeschla EM. Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. *Journal of molecular biology* 2006; **360**(4)**:** 760-773.

23. Shun MC, Botbol Y, Li X, Di Nunzio F, Daigle JE, Yan N *et al.* Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human immunodeficiency virus type 1 infectivity. *Journal of virology* 2008; **82**(23)**:** 11555-11567.

24. Meehan AM, Saenz DT, Morrison JH, Garcia-Rivera JA, Peretz M, Llano M *et al.* LEDGF/p75 proteins with alternative chromatin tethers are functional HIV-1 cofactors. *PLoS pathogens* 2009; **5**(7)**:** e1000522.

25. Ciuffi A, Diamond TL, Hwang Y, Marshall HM, Bushman FD. Modulating target site selection during human immunodeficiency virus DNA integration in vitro with an engineered tethering factor. *Human gene therapy* 2006; **17**(9)**:** 960-967.

26. Sun D, Melegari M, Sridhar S, Rogler CE, Zhu L. Multi-miRNA hairpin method that improves gene knockdown efficiency and provides linked multi-gene knockdown. *BioTechniques* 2006; **41**(1)**:** 59-63.

27. Bustin M, Catez F, Lim JH. The dynamics of histone H1 function in chromatin. *Molecular cell* 2005; **17**(5)**:** 617-620.

28. Catez F, Ueda T, Bustin M. Determinants of histone H1 mobility and chromatin binding in living cells. *Nature structural & molecular biology* 2006; **13**(4)**:** 305-310.

29. Maison C, Almouzni G. HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol* 2004; **5**(4)**:** 296-304.

30. Ochs RL, Muro Y, Si Y, Ge H, Chan EK, Tan EM. Autoantibodies to DFS 70 kd/transcription coactivator p75 in atopic dermatitis and other conditions. *The Journal of allergy and clinical immunology* 2000; **105**(6 Pt 1)**:** 1211-1220.

31. Llano M, Delgado S, Vanegas M, Poeschla EM. Lens epithelium-derived growth factor/p75 prevents proteasomal degradation of HIV-1 integrase. *The Journal of biological chemistry* 2004; **279**(53)**:** 55570-55577.

32. Minc E, Allory Y, Courvalin JC, Buendia B. Immunolocalization of HP1 proteins in metaphasic mammalian chromosomes. *Methods Cell Sci* 2001; **23**(1-3)**:** 171-174.

33. Hayakawa T, Haraguchi T, Masumoto H, Hiraoka Y. Cell cycle behavior of human HP1 subtypes: distinct molecular domains of HP1 are required for their centromeric localization during interphase and metaphase. *Journal of cell science* 2003; **116**(Pt 16)**:** 3327-3338.

34. Ibrahimi A, Vande Velde G, Reumers V, Toelen J, Thiry I, Vandeputte C *et al.* Highly efficient multicistronic lentiviral vectors with peptide 2A sequences. *Human gene therapy* 2009; **20**(8)**:** 845-860.

35. Hacker CV, Vink CA, Wardell TW, Lee S, Treasure P, Kingsman SM *et al.* The integration profile of EIAV-based vectors. *Mol Ther* 2006; **14**(4)**:** 536-545.

36. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC *et al.* Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS biology* 2004; **2**(8)**:** E234.

37. Berry C, Hannenhalli S, Leipzig J, Bushman FD. Selection of target sites for mobile DNA integration in the human genome. *PLoS computational biology* 2006; **2**(11)**:** e157.

38. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 2001; **410**(6824)**:** 116-120.

39. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 2001; **410**(6824)**:** 120-124.

40. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* 2007; **129**(4)**:** 823-837.

41. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics* 2008; **40**(7)**:** 897-903.

42. Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & development* 2004; **18**(11)**:** 1251-1262.

43. Jones B, Su H, Bhat A, Lei H, Bajko J, Hevi S *et al.* The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. *PLoS genetics* 2008; **4**(9)**:** e1000190.

44. Vogel MJ, Guelen L, de Wit E, Peric-Hupkes D, Loden M, Talhout W *et al.* Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome research* 2006; **16**(12)**:** 1493-1504.

45. Bushman FD. Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. *Proceedings of the National Academy of Sciences of the United States of America* 1994; **91**(20)**:** 9233-9237.

46. Holmes-Son ML, Chow SA. Correct integration mediated by integrase-LexA fusion proteins incorporated into HIV-1. *Mol Ther* 2002; **5**(4)**:** 360-370.

47. Bushman FD, Miller MD. Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. *Journal of virology* 1997; **71**(1)**:** 458-464.

48. Tan W, Dong Z, Wilkinson TA, Barbas CF, 3rd, Chow SA. Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. *Journal of virology* 2006; **80**(4)**:** 1939-1948.

49. Hare S, Shun MC, Gupta SS, Valkov E, Engelman A, Cherepanov P. A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75. *PLoS pathogens* 2009; **5**(1)**:** e1000259.

50. Geraerts M, Michiels M, Baekelandt V, Debyser Z, Gijsbers R. Upscaling of lentiviral vector production by tangential flow filtration. *The journal of gene medicine* 2005; **7**(10)**:** 1299-1310.

51. Girton JR, Johansen KM. (2008) Chromatin structure and the regulation of gene expression: The lessons of PEV in drosophila. Adv Genet 61: 1-43.

52. Lewinski M, Bisgrove D, Shinn P, Chen H, Verdin E, et al. (2005) Genome-wide analysis of chromosomal features repressing HIV transcription. J. Virol. 79, 6610-6619.

53. Garcia-Rivera JA, Bueno MT, Morales E, Kugelman JR, Rodriguez DF, et al. (2010) Implication of serine residues 271, 273, and 275 in the human immunodeficiency virus type 1 cofactor activity of lens epithelium-derived growth factor/p75. J Virol 84(2): 740-752.

CHAPTER 4 – INTEGRATION SITE DISTRIBUTION IN MICE FOLLOWING THERAPEUTIC GENE TRANSFER TO TREAT β-THALASSEMIA

The contents of this chapter are being prepared for submission:

Ronen K, Negre O, Malani N, Denaro M, Gillet-Legrand B, Leboulch P, Down JD, Bushman FD. Integration site distribution in mice following therapeutic gene transfer to treat beta-thalassemia.

4.1 Abstract

Treatment of genetic diseases such as β-thalassemia (β-Thal) requires covalent integration of therapeutic genes into a patient's chromosome to allow stable inheritance. The Lentiglobin™ lentiviral vector has been applied to gene therapy for β-Thal with success in one human patient, but a semi-dominant clonal expansion after integration in the HMGA2 locus in this patient raised the question of whether lentiviral integration could alter activity of nearby genes and promote abnormal cellular growth. Here we have used a mouse model for therapeutic gene transfer and bone marrow transplantation, using the same vector used to treat β-thal patients, and investigated the integration site distributions present after 9 months of hematopoietic reconstitution in five busulfan pre-treated β-Thal mice. The recipient mice demonstrated correction of the disease and were healthy at time of sacrifice. The pre-transplantation integration site distribution was typical of lentiviral vectors, showing favored integration in genes and gene-rich regions. After hematopoeitic cell repopulation in mice, integration sites located near genes involved in growth control

were not enriched. No integration sites in or near HMGA2 were detected. Cells containing integration sites in genes became less common after growth in mice than before transplantation, and this was accentuated after subsequent culture of explanted cells in methylcellulose. This is consistent with selective loss of cells containing integration sites in genes, possibly due to changes in dosage. Similar results have been seen in some but not all previous studies. Most importantly, these data in mice indicate that gene correction can be achieved without any indication of vector-enhanced cell proliferation.

4.2 Introduction

Retroviral vectors have been successfully used in human trials of gene transfer to treat a number of genetic diseases, including X-linked severe combined immunodeficiency disorder (X-SCID) [1], adenosine deaminase deficiency (ADA-SCID) [2], chronic granulomatous disease (CGD) [3] and X-linked adrenoleukodystrophy (ALD) [4]. However, several adverse events have occurred in which integration of the therapeutic vector resulted in insertional activation of proto-oncogenes, contributing to the development of leukemia [5-9]. Thus there has been intense interest in characterizing the integration profile of gene therapy vectors and improving their safety.

While the majority of completed trials of retroviral gene transfer have used gammaretroviral vectors, lentiviral vectors are increasingly used. Their appeal stems from several observations. Lentiviral vectors, unlike gammaretroviral vectors, infect non-dividing cells [10]. Additionally, no convincing examples of insertional

activation of oncogenes and consequent transformation have been reported associated with HIV infection, though HIV proviruses can affect the activity of nearby genes [11, 12]. Moreover, studies in tumor-prone mouse models have reported less genotoxicity resulting from lentiviral than gammaretroviral transduction [13, 14]. One possible explanation for the difference in oncogenic potential between lentiviral and gammaretroviral vectors may be related to differences in their preferred sites of integration in the genome. Insertional activation in animal models and human patients is caused by integrated vector promoter or enhancer elements upregulating downstream cellular genes [5, 6, 8, 15-19]. Gammaretroviral vectors show a strong propensity to integrate at promoters and gene 5' ends [20], and clustering near genes controlling cell growth and proliferation has been reported [21]. Lentiviruses, on the other hand, favor integration in the bodies of transcription units, avoiding regulatory 5' regions [22-24], with no bias for growth-associated genes. The lentiviral integration pattern may thus be less likely to result to insertional activation, though other variables such as cell type specificity may also play a role [25]. In the context of vectors, the engineered transcriptional control elements may also explain the observed differences between gammaretroviral and lentviral vectors [13, 14].

Three clinical trials have been conducted using lentiviral vectors in humans, and in each the genomic distribution of integration sites was monitored. The first tested an anti-HIV therapy by delivering an antisense HIV-*env* gene to mature T-cells [26]. Integration events in these patients showed no evidence of enrichment of sites in proto-oncogenes following transduction [27]. The second trial treated two patients

with ALD by transduction of hematopoietic stem cells [4]. Integration sites in these two patients also showed sustained polyclonality up to 24 months after transplantation. Thirdly, one patient was treated for β-thalassemia by hematopoietic stem cell transduction with a lentiviral vector encoding β-globin with clinical success (Cavazzana-Calvo *et al.*, submitted). In the β-Thal trial, the integration site distribution determined at 19 months post-transplantation showed that ~50% of integration sites were within the proto-oncogene HMGA2. The integration event was associated with increased transcription of HMGA2 and expression of a transcript whose 3' UTR was replaced by vector sequences. This 3'-subsituted transcript lacked the target of the repressive miRNA, *let-7*, contributing to elevated HMGA2 expression along with an increase in the rate of transcriptional initiation. HMGA2 has been implicated as important in persistence of stem cells [28] and is disregulated in some cancers, including by disruption of the normal gene 3' end [29-32]. While the treated patient remains healthy, this finding has raised questions about the possibility of lentiviral integration in or near growth-control genes imparting a selective advantage and leading to preferential outgrowth of the gene-modified cell.

The same SIN lentiviral vector as was used in the β-Thal clinical trial has been tested in a preclinical mouse model of β-thalassemia, using a closely related transduction protocol (Negre *et al.*, submitted). This allows further study of the vector's possible genotoxicity, and a comparison of the distribution of integration sites in the mouse and human studies. Nine months post-transplantation the mice

showed no apparent pathological abnormalities while demonstrating long-term resolution of the hematological abnormalities associated with β-thalassemia.

Here we present a study of the vector integration sites recovered from bone marrow transduction and transplantation of five mice. We observed oligoclonality of integration sites in all animals. Functional classes of genes close to integration events showed no significant enrichment in genes encoding growth-related functions. Integration events with the potential to disrupt onocogene regulation by disruption of miRNA regulation through the mRNA 3' end were not detectably enriched. We also observed that integration sites from cells following growth either in mice or in subsequent cell culture showed a reduced frequency in transcription units and gene-dense regions, suggestive of selection against integration events within genes during long term passage.

4.3 Material and Methods

Vector transduction and preparation of bone marrow samples

Vector design and transduction is described in (Negre *et al.*, submitted). Briefly, clinical grade VSV-G pseudotyped lentiviral supernatant was produced by transient transfection of HEK293T cells with a 5-plasmid system. Bone marrow cells were isolated from three-month-old female β-thalassemic mice, injected with 5-fluoro-uracil (150 mg/kg) 4 days previously. Nucleated cells were isolated and transduced for 24h, after which they were washed and resuspended in PBS. 400,000 cells were injected via the retro-orbital sinus into each of five 6-month old recipient male β-thalassemic mice pre-treated with 4 daily doses of 20 mg/kg busulfan.

500,000 transduced cells from the transplant inoculum were grown in liquid culture for 11 days, and 30,000 plated in methylcellulose and cultured for 7 days in triplicate. Following culture, methylcellulose was dissolved, colonies harvested and genomic DNA extracted. Recipient mice were sacrificed 9.2 months after transplantation and bone marrow harvested. 3 million cells from each mouse were used for immediate genomic DNA extraction, and 90,000 cells from each mouse were cultured in methylcellulose for 7 days before genomic DNA extraction.

Isolation of integration site sequences

Integration site isolation was performed by ligation-mediated PCR essentially as described previously [22, 24, 33, 34]. Each DNA sample (420-1000ng) was digested with MseI and NlaIII separately. Linkers were ligated to the digested samples and samples treated with ApoI to limit amplification of the internal vector fragment downstream of the 5' LTR. Samples were then amplified by nested PCR and sequenced by 454 pyrosequencing at the University of Pennsylvania DNA sequencing center. In order to sequence all amplicons in one sequencing run, PCR primers contained 8bp barcodes between the 454 sequencing primer and the region complementary to the LTR.

Bioinformatic analysis

Integration sites were determined to be authentic if the sequences began within 3bp of vector LTR ends, had a >98% match to the mouse genome (mm8 draft), and had a unique best hit when aligned to the mouse genome by BLAT. All integration sequences will be deposited in GenBank upon acceptance of the paper for

publication. Integration sites can be viewed on the UCSC browser at the following URL (http://genome.ucsc.edu/cgi-bin/hgTracks?db=mm8&hgt.customText=http://microb230.med.upenn.edu/ucsc/Rone n-BetaThalSites.bed.gz). Three integration sites were found in sequences barcoded as more than one mouse and deduced to be likely due to crossover between samples during PCR (though transduced cells were pooled before transplantation, providing another candidate explanation). The probable origin of each site was assigned based on sequence abundance and recovery in multiple samples from the same mouse, and excluded from analysis of the other mouse from which it was recovered. For each experimental integration site three matched random control sites were computationally generated. These sites were matched to the experimental sites in their distance to the nearest MseI or NlaIII site as appropriate.

The RTCGD cancer gene database is available at http://rtcgd.ncifcrf.gov/. The expanded allOnco cancer genes list is described at http://microb230.med.upenn.edu/protocols/cancergenes.html.

Enrichment of integration events relative to various genomic features was compared between datasets by Fisher's exact (where stated) or by multiple regression models for integration intensity and a c-logit test for significance, as described in Berry *et al*. [35]. Receiver Operating Characteristic (ROC) analysis was used to represent the relationship between integration and various genomic features by a single numerical value [35], which was used to generate a colored heat map. Genomic features analyzed in Figure 4-3 are as follows. 'In Refseq' shows the

preference for integration within genes called by the Refseq gene call. 'Gene width' shows the relationship with the length of the gene harboring the integration site. 'Intergene width' shows the relationship with the length of the interval between genes. Short genes and short intergene lengths are associated with gene-rich regions. 'Gene start distance' shows the relationship to the distance to the nearest gene 5' end. 'Gene boundary distance' to the nearest gene 5' or 3' end. 'Refseq count' shows the relationship to the number of Refseq genes within a given window around each integration site (windows shown as 1Mb, 100kb etc). 'Expression' is gene activity, measured in MEFs using Affymetrix microarrays. Genes were ranked for relative expression and the relationship is shown between integration and the expression level over given windows around ech site. 'Top 1/2' and 'Top 1/16' means only genes in the upper half or sixteenth were scored, respectively. 'CpG count' is calculated analogously to 'Refseq count', counting the number of CpG islands in the window specified. 'CpG desnsity' takes the number of CpG islands in the given window and divides by the number of base-pairs for a density measure. 'GC content' denotes the percentage G/C residues in the sequence surrounding each integration site, in the window shown. Analysis was carried out in R (http://www.r-project.org).

Ingenuity Pathway Analysis software (http://www.ingenuity.com) was used to form networks based on gene lists assembled from the gene nearest to each integration site (Figure 4-2).

miRNA predictions were annotated based on three online tools: miRbase (http://microrna.sanger.ac.uk/sequences/), TargetScan

(http://www.targetscan.org/mmu_50/) and miRDB

(http://mirdb.org/miRDB/index.html)

4.4 Results

Isolation of integration sites from Lentiglobin-transduced bone marrow before and after transplantation

Nucleated cells isolated from 5-fluorouracil-treated 3-month-old β-thalassemic mice were transduced with the Lentiglobin vector for 24 hours. Cells were harvested at this time for integration site analysis (pre-transplantation samples). Before isolating genomic DNA, cells were subjected to liquid culture or methylcellulose culture. Methylcellulose culture selects for committed hematopoietic progenitor cells, which have sufficient replicative capacity to form colonies under the *ex vivo* culture conditions. Culturing cells in this way therefore allowed us to study the integration sites of progenitor cells, and compare them to sites from total bone marrow, which also includes terminally differentiated cells. Colonies derived from these progenitors were then isolated from the culture and genomic DNA extracted. 500,000 pre-transplantation cells were cultured in liquid media for 11 days; 90,000 in methylcellulose-based media for 7 days, yielding about 2000 colonies from myeloid progenitors. Nine months after transplantation, bone marrow was also isolated for integration site analysis (post-transplantation). Post-transplantation cells were either cultured in methylcellulose as above, or DNA was extracted immediately. Three million bone marrow cells were used for immediate DNA extraction, and 90,000 bone

marrow cells per mouse were cultured in methylcellulose, yielding about 50 colonies from myeloid progenitors per mouse.

Vector integration sites were amplified from the genomic DNA as previously described [17]. Since this method relies on restriction digestion, which has been shown to introduce a significant recovery bias [17], samples were processed with two different restriction enzymes in parallel, MseI and NlaIII, in an attempt to maximize site recovery. Vector-host junctions were sequenced by 454/Roche pyrosequencing, using 8bp barcodes in the primer to distinguish between sets run simultaneously.

Table 4-1 summarizes the number of integration sites identified, pooling across the two restriction enzymes used on each sample. Across all sets, we obtained a total of 31128 integrations passing quality control, representing 1162 unique sites over all sets. The numbers of unique sites detected in each mouse ranged from 25-28, providing an initial measure of the numbers of gene-corrected long-term repopulating cell clones present.

However, the methods used for integration site recovery are unlikely to capture all integration sites, so we investigated methods for estimating the size of the full population. Since two different enzymes were used to isolate integration sites from each sample, overlap between sites recovered by the two enzymes can be used to estimate the total number of sites using a capture-recapture approach. Using the Lincoln-Petersen method [36], the number of unique sites estimated to be present in each mouse are 58, 48, 75, 140 and 48 for mice 31, 32, 33.1, 33.2 and 34 respectively (see supplementary table S4-1). In another approach, based on a previous study

Data Set	Total reads	Unique sites	Vector	Description	Reference
Pre-transplantation (liquid)	5254	631	Lentiglobin	BM pre-transplantation (liquid culture 11 days)	This study
Pre-transplantation (methylcellulose)	5430	338	Lentiglobin	BM pre-transplantation (methyl-cellulose culture 7 days)	This study
Pre-transplantation total	10684	963	Lentiglobin	BM pre-transplantation (all sites pooled)	This study
Mouse 31 (uncultured)	4209	38	Lentiglobin	BM 9 months post-transplantation - mouse 31	This study
Mouse 32 (uncultured)	709	28	Lentiglobin	BM 9 months post-transplantation - mouse 32	This study
Mouse 33.1 (uncultured)	4458	34	Lentiglobin	BM 9 months post-transplantation - mouse 33.1	This study
Mouse 33.2 (uncultured)	1105	25	Lentiglobin	BM 9 months post-transplantation - mouse 33.2	This study
Mouse 34 (uncultured)	4696	32	Lentiglobin	BM 9 months post-transplantation - mouse 34	This study
Post-transplantation (uncultured)	15177	158	Lentiglobin	BM 9 months post-transplantation - 5 mice pooled	This study
Mouse 31 methyl-cellulose (methylcellulose)	314	5	Lentiglobin	BM 9 months post-transplantation (methyl-cellulose culture 7 days) - mouse 31	This study
Mouse 32 methyl-cellulose (methylcellulose)	1302	7	Lentiglobin	BM 9 months post-transplantation (methyl-cellulose culture 7 days) - mouse 32	This study
Mouse 33.1 methyl-cellulose (methylcellulose)	1270	19	Lentiglobin	BM 9 months post-transplantation (methyl-cellulose culture 7 days) - mouse 33.1	This study
Mouse 33.2 methyl-cellulose (methylcellulose)	759	43	Lentiglobin	BM 9 months post-transplantation (methyl-cellulose culture 7 days) - mouse 33.2	This study
Mouse 34 methyl-cellulose (methylcellulose)	1622	8	Lentiglobin	BM 9 months post-transplantation (methyl-cellulose culture 7 days) - mouse 34	This study
Post-transplantation (methylcellulose)	5267	81	Lentiglobin	BM 9 months post-transplantation (methyl-cellulose culture 7 days) - 5 mice pooled	This study
Post-transplantation total	20444	201	Lentiglobin	BM 9 months post-transplantation (all sites pooled)	This study
Mouse Lenti Tumor	3353	225	Lentiviral	Various tissues from mice treated with lentiviral vectors that developed tumors	[13]
HIV MEF	3239	2532	SIN HIV-GFP	Murine embryonic fibroblasts infected in culture	[13]

Table 4-1. Integration site data sets used in this study.

where six restriction enzymes were independently used to isolate integration sites from patients treated for X-SCID (Wang *et al.*, submitted) we asked what proportion of all sites was recovered by MseI alone or NlaIII alone. Based on that study, approximately 25% of all unique sites were isolated using either enzyme alone (data not shown). According to this estimate, the numbers of unique sites in Table 4-1, which were isolated using both MseI and NlaIII, likely represent ~50% of the total sites present in the mice.

The number of sites recovered per mouse was lower after methylcellulose culture than in samples that were not cultured in methylcellulose. Ninety-thousand cells from each mouse were placed into methylcellulose culture, forming ~50 colonies from which DNA was extracted, and 3 million cells were taken for immediate DNA extraction. Evidently plating and growth in methylcellulose selected out a subset of progenitor cells.

Figure 4-1 shows the proportion of the most abundant integration sites identified in the bone marrow of each transplanted mouse, comparing sites from cells with and without methylcellulose culture. Each site is classified by the gene it is either within or closest to, and the abundance of each site is displayed. It can be seen that integration events post-transplantation were oligoclonal, with levels of clonal dominance varying from 40% to 70% in different mice.

Integration is not enriched near genes associated with growth or oncogenesis

We next investigated the functional categories of the genes close to integration sites. Insertional oncogenesis by retroviral vectors most commonly results from

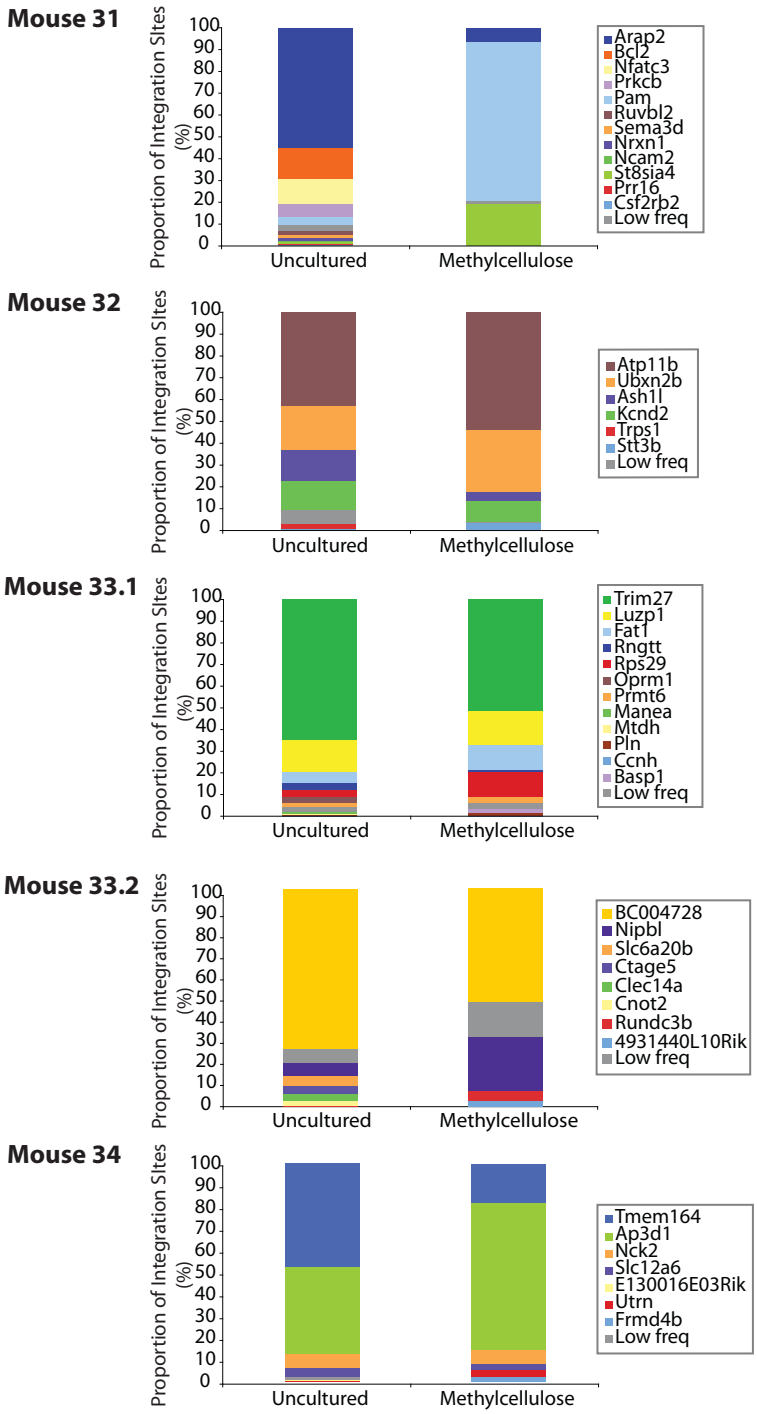

Figure 4-1. Integration site frequencies in individual mice, 9 months after Lentiglobin-transduced bone marrow transplantation. Each integration site is labeled by the gene it falls within or the nearest gene, and represented by a different color. The proportions of sequences recovered of each site are shown. Sites recovered fewer than ten times were pooled and are displayed as 'Low Freq'.

activation of proto-oncogenes or, less commonly, inactivation of tumor-suppressors following nearby integration. Thus, cells harboring integration sites close to growth-control genes can have a selective advantage and become enriched *in vivo*. We therefore asked whether integration sites post-transplantation showed evidence of such enrichment near genes associated with cell growth. Table 4-2 shows the proportion of unique integration sites that lie within 50kb of a cancer gene, as defined by studies of insertional activation in mice (the RTCGD [37]). Integration sites from all 5 mice were pooled in this analysis. The proportion of sites in each data set is compared to the proportion pre-transplantation. No statistically significant differences between sets were found by the Fisher's exact test for proximity to cancer genes. We also repeated these analyses with an extended list of cancer-related genes (the allOnco data set from http://microb230.med.upenn.edu/protocols/cancergenes.html). This list is a compilation of several gene lists of cancer-associated genes from diverse vertebrates, in which all genes in any organism were mapped to their murine homologs (see Materials and Methods). Results with this expanded list also did not show any statistically significant differences between sets (data not shown).

A related question centered on whether the proportion of sites in each dataset where the closest gene was an RTCGD gene was increased after growth of cells in mice. Integration sites were pooled over all mice for this analysis. These values also did not differ between groups, or from random (Table 4-2).

Integration Set	Sites in genes	Sites in exons	Percent in exons	Sites in same orientation as transcription	Percent same orientation as transcription
Pre-transplantation (liquid)	427	33	7.73	228	53.40
Pre-transplantation (methylcellulose)	219	13	5.94	112	51.14
Pre-transplantation (all)	641	46	7.18	336	52.42
Post-transplantation (uncultured)	83	5	6.02	39	46.99
Post-transplantation (methylcellulose)	23	0	0.00	13	56.52
Post-transplantation (all)	89	5	5.62	41	46.07
HIV MEFs	1745	106	6.07	860	49.28
Mouse lenti tumors	138	10	7.25	77	55.80
MRC pre-transplantation (liquid)	632	34	5.38	301	47.63
MRC pre-transplantation (methylcellulose)	370	23	6.22	193	52.16
MRC pre-transplantation (all)	997	56	5.62	490	49.15
MRC post-transplantation (uncultured)	172	4	2.33	78	45.35
MRC post-transplantation (methylcellulose)	75	4	5.33	35	46.67
MRC post-transplantation (all)	203	7	3.45	93	45.81
MRC HIV MEF	2492	127	5.10	1273	51.08
MRC mouse lenti tumors	247	13	5.26	133	53.85

Table 4-2. Integration site frequency in the vicinity of oncogenes.

The above analysis did not take into account the frequency of recovery of different integration sites. It could be imagined that an early indication of insertional activation would be the overrepresentation of integration sites close to genes involved in growth control. We therefore asked whether there was a correlation between the frequency of recovery of a particular site and its proximity to genes in the RTCGD. Integration sites from each set were grouped into 3 bins of increasing recovery frequency, then the proportion of sites in each bin that fell within 50kb of an RTCGD calculated. No significant correlations between frequency of recovery and proximity to cancer genes were found in any data set (data not shown). Thus these data provide no evidence of clonal expansion of cells bearing integration sites near growth control genes.

We also used the Ingenuity network analysis software to study the functional categories of the genes close to integration sites. Clustering of genes into networks related to growth control after growth in mice would be suggestive of selection for cells whose growth has been activated by vector integration. Post-transplantation sites were pooled across mice. The highest-scoring networks based on integration sites pre-transplantation and post-transplantation are shown in Figure 4-2. It can be seen that in both cases, genes with functions relating to growth control and cell death are represented, but growth in mice does not appear to have selected differentiallly for cells bearing integration sites near such genes.

In the human β-Thal trial (Cavazzana-Calvo *et al.*, submitted) outgrowth was detected of an integration site in the HMGA2 gene. This was in the sense orientation

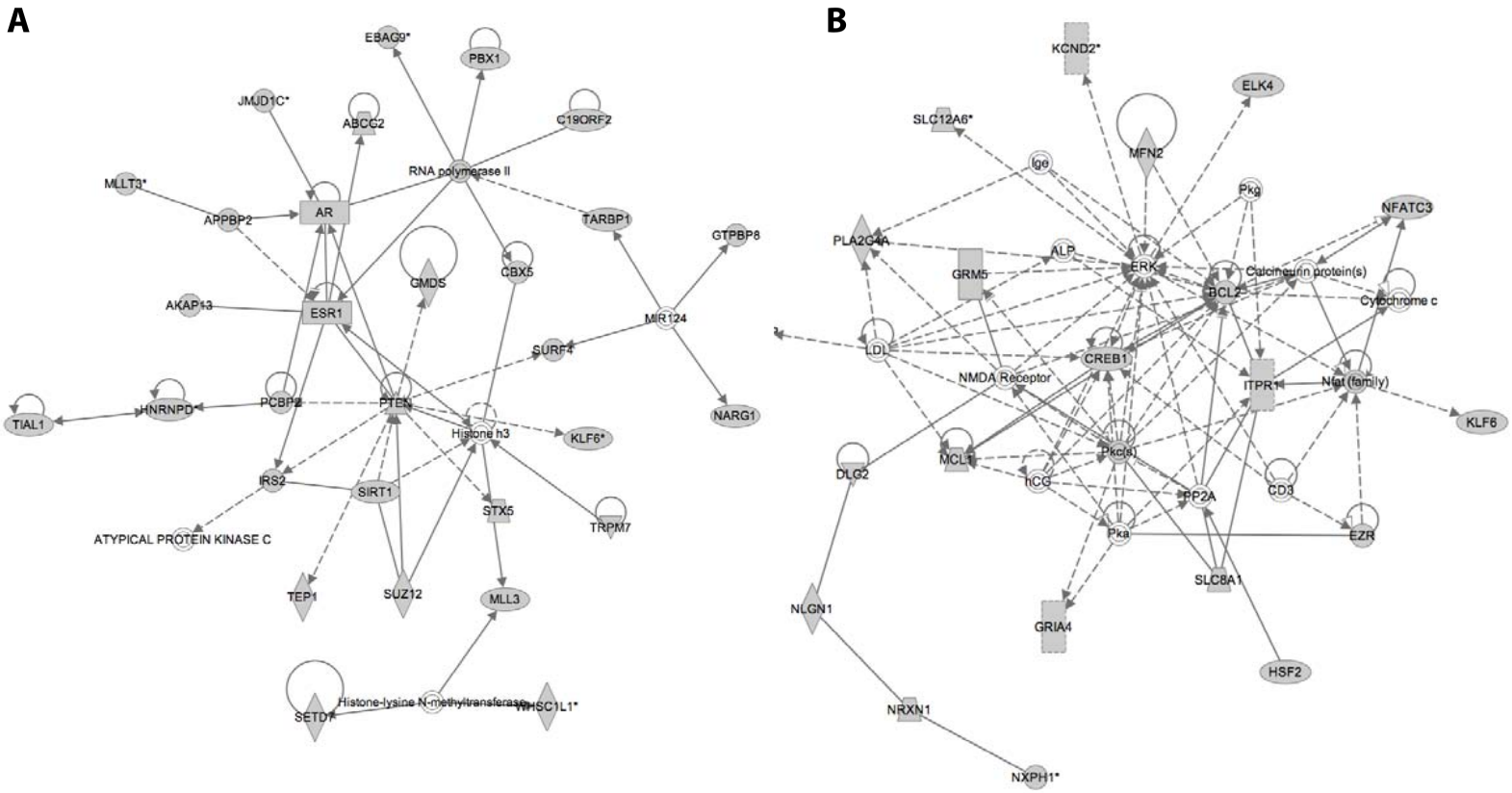

Figure 4-2. Network analysis of genes near integration sites before and after transplantation. The highest-scoring networks generated by the Ingenuity software from the genes closest to integration sites are shown. A. pre-tranplsantation, network 'Cancer, Neurological Disease, Hematological Disease', 30 focus molecules. B. post-transplantation, network 'Cell Cycle, Skeletal and Muscular System Development and Function, Cell Death', 21 focus molecules. Gene names on gray backgrounds are genes near integration sites; gene names on white backgrounds represent interactors added to the network computationally to enhance connectivity. The number of genes from the input list used in each network are listed ('focus molecules'). Solid lines represent direct interactions, dotted lines indirect.

in the third intron of the gene, upstream of miRNA binding sites in the 3'UTR, leading to 3' end substitution in the HMGA2 mRNA and potential release from post-transcriptional repression. We thus examined the most abundant sites in each mouse looking for outgrowth of integration sites lying in the sense orientation in the intron of a gene with predicted miRNA binding in the 3' UTR. Of the seven most abundant sites from five mice under the two culture conditions (Arap2 and Pam in mouse 31, uncultured and methylcellulose cultured respectively, Atp11b in mouse 32, Trim27 in mouse 33.1, BC004728 in mouse 33.2, Tmem164 and Ap3d1 in mouse 34), two sites (Pam and Tmem164) had these properties. Neither of these genes has been associated with oncogenesis. Three of the seven were in the antisense orientation in introns, with miRNAs predicted in their 3' UTR (Atp11b, BC004728, Ap3d1). BC004728 has been associated with metastasis [38, 39], but not tumorigenesis. By comparison, of 963 unique pre-transplantation sites, 311 were in the sense orientation in an intron and 284 antisense in an intron. We were thus unable to find any strong evidence associating clonal expansion with vector integration in the sense orientation within a growth control gene subject to miRNA regulation.

Integration sites after growth of transduced cells *in vivo* or in methylcellulose culture are less frequently distributed in transcription units

Lentiviral vectors show a preference for integration in particular genomic features, such as the bodies of transcription units and gene-rich regions of the genome [20, 22, 23, 35]. Figure 4-3 shows a heatmap representing the genomic distribution of integration sites in four types of samples – pre-transplantation and post-

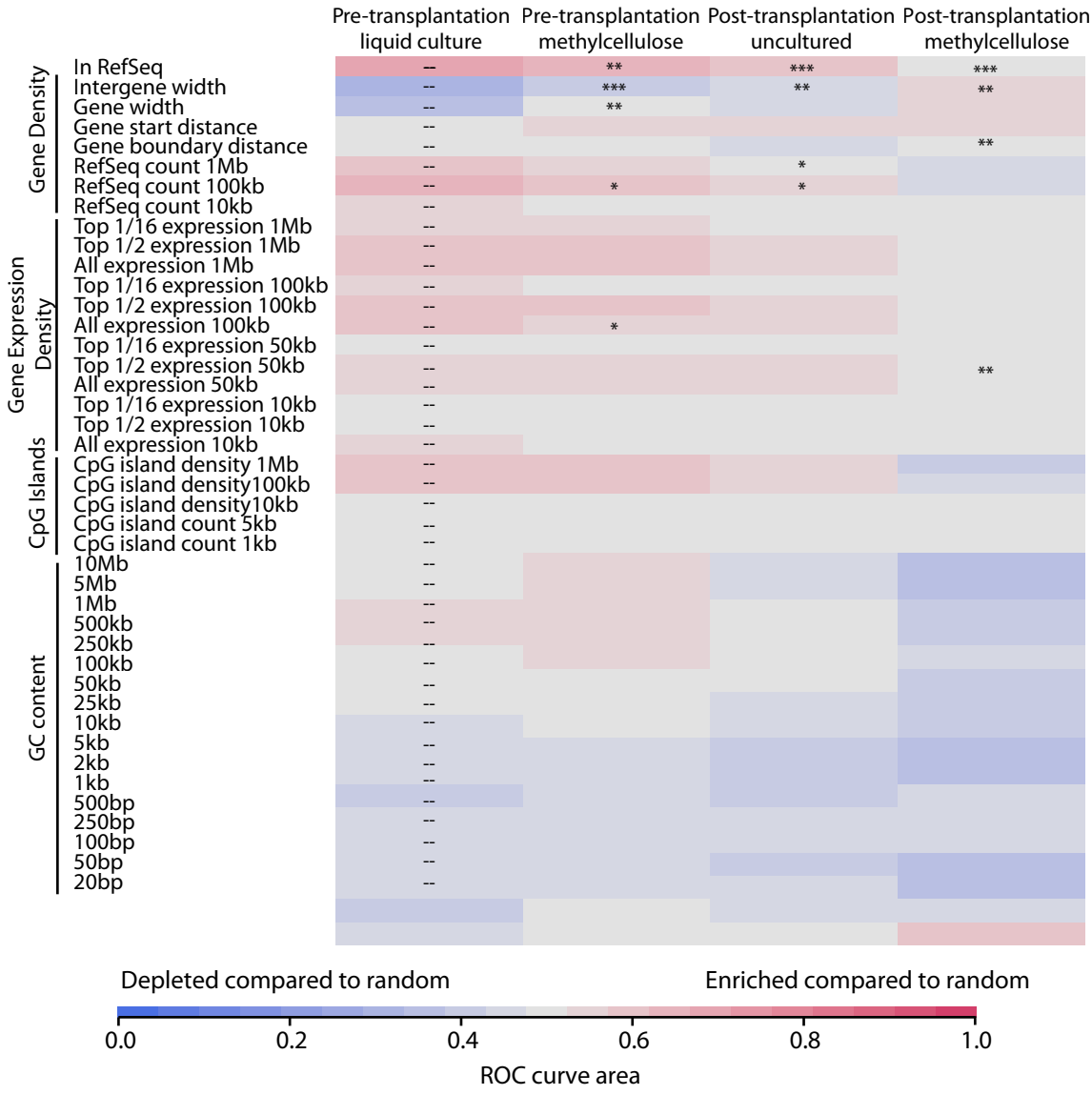

Figure 4-3. Heat map illustrating genomic distribution of integration sites. Favoring or disfavoring of a genomic feature within a window around integration sites in each data set is represented as a colored tile. The color is determined by ROC curve area comparing the density of the feature near experimental sites and matched random control sites. See Materials and Methods for explanation of genomic features. The p-value for the comparison with pre-transplantation liquid culture, determined by a logistic regression method that respects the pairing in the data (clogit), is overlaid on the heatmap tile (*p<0.05; **p<0.01; ***p<0.001).

transplantation, each with and without methylcellulose culture. For this analysis integration sites from all mice were pooled within a sample type. For each genomic feature, favoring or disfavoring of integration in the feature compared to random is represented by red or blue coloring respectively. The integration site distribution pre-transplantation was consistent with previously published studies, with a favoring of transcription units and regions with high gene density, narrow genes, and high CpG island density. However, post-transplantation and particularly following methylcellulose culture these trends weakened and in some cases reversed (asterisks represent statistically significant deviations from the pre-transplantation liquid culture dataset).

Figure 4-4 shows a graphical representation of the frequency of integration in transcription units in each set. It can be seen that while integration pre-transplantation is significantly enriched in RefSeq transcription units, cell growth both in mice and in methylcellulose decreased the preference for provirus accumulation in transcription units. 67.7% of integration sites pre-transplantation without culture were found in transcription units. In pre-transplantation cells cultured in methylcellulose, 64.8% of sites were in transcription units ($p=0.0097$ for the difference between the two). Integration in post-transplantation cells in liquid culture was 52.5% in genes ($p=0.014$ for the difference with pre-transplantation uncultured cells) and in post-transplantation cells grown in methylcellulose culture 27.7% ($p=1.59e-6$ for the difference with pre-transplantation uncultured cells). Integration in transcription units is thus below the level expected by chance after growth in mice

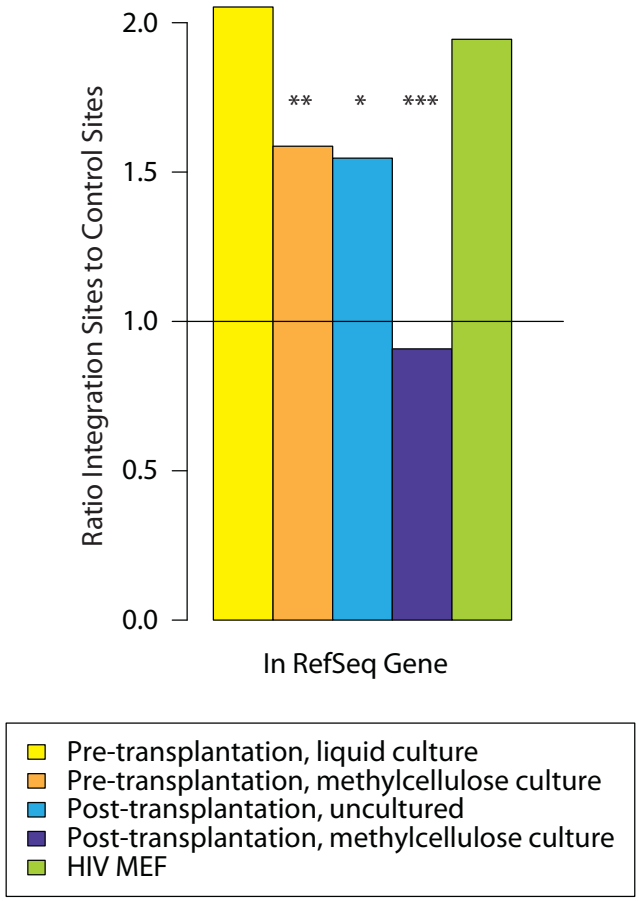

Figure 4-4. Integration in transcription units. The proportion of integration sites in each data set within or outside of transcription units are shown, normalized to matched random control sites (indicated by the horizontal line). Significant differences from pre-transplantation (liquid culture) sites is denoted by asterisks (*p<0.05; **p<0.01; ***p<0.001).

and culture in methylcellulose, suggestive of selection against cells with integrated proviruses disrupting genes sensitive to dosage changes.

We therefore asked whether any bias in the orientation of the provirus relative to transcription or integration in introns versus exons could be seen, since proviruses in different orientations may have different effects on host gene activity. No statistically significant effects were seen (Table 4-3). We thus conclude that provirus accumulation within transcription units was lower after prolonged growth, but that this was not associated with a bias in provirus orientation relative to the host transcription unit.

4.5 Discussion

The safety of lentiviral vectors for gene delivery has been a matter of intense interest in the gene therapy field, with extensive resources invested in optimizing vector design to minimize the risk of genotoxicity. Trials of gammaretroviral vectors in humans have led to adverse clinical events associated with clonal enrichment of integration sites near oncogenes and tumor-suppressors, but the three human trials of lentiviral gene therapy have not led to clinical adverse events to date. In the third trial, involving transduction of hematopoietic stem cells for the treatment of β-thalassemia, relative clonal dominance of cells containing a site in the proto-oncogene HMGA2 was observed (Cavazzana-Calvo *et al.*, submitted), though the patient remains healthy. Here, we present a study of integration by the lentiviral vector from this study in a mouse model of β-thalassemia.

Integration Set	Sites in genes	Sites in exons	Percent in exons	Sites in same orientation as transcription	Percent same orientation as transcription
Pre-transplantation (liquid)	427	33	7.73	228	53.40
Pre-transplantation (methylcellulose)	219	13	5.94	112	51.14
Pre-transplantation (all)	641	46	7.18	336	52.42
Post-transplantation (uncultured)	83	5	6.02	39	46.99
Post-transplantation (methylcellulose)	23	0	0.00	13	56.52
Post-transplantation (all)	89	5	5.62	41	46.07
HIV MEFs	1745	106	6.07	860	49.28
Mouse lenti tumors	138	10	7.25	77	55.80
MRC pre-transplantation (liquid)	632	34	5.38	301	47.63
MRC pre-transplantation (methylcellulose)	370	23	6.22	193	52.16
MRC pre-transplantation (all)	997	56	5.62	490	49.15
MRC post-transplantation (uncultured)	172	4	2.33	78	45.35
MRC post-transplantation (methylcellulose)	75	4	5.33	35	46.67
MRC post-transplantation (all)	203	7	3.45	93	45.81
MRC HIV MEF	2492	127	5.10	1273	51.08
MRC mouse lenti tumors	247	13	5.26	133	53.85

Table 4-3. Integration in gene features.

We found that 9 months after transplantation integration sites in the bone marrow were oligoclonal. The genes closest to integration sites did not show evidence of enrichment in growth-related functional categories relative to genes targeted in pre-transplantation sites. Similarly, no enrichment of integration sites within 50kb of proto-oncogenes was found post-transplantation, and no relationship between the proximity of integration sites to oncogenes and site abundance.

We estimated the numbers of gene-corrected progenitor cells in each mouse to be roughly 50-150 based on the numbers of unique integration sites recovered and upward corrections to account for sparse sampling. It is of interest to compare these numbers to those expected based on estimates of stem cell proportions in bone marrow. A total of 400,000 transduced cells were transplanted into each mouse. It can be estimated that following treatment with 5-fluorouracil mouse bone marrow contains 1 in 10,000 stem cells [40-42]. In the transduction conducted in this study, an average of 70% of cells were transduced (Negre *et al*, submitted), so recovery of about 28 clones would be expected on average per mouse. The numbers of clones per mouse estimated from vector marking were somewhat higher, ranging from ~50-150 clones. We note that biases in recovery due to use of restriction enzyme cleavage would tend to inflate the estimates by artificially reducing the overlap between sets, possibly in part accounting for the differences. However, the estimates from vector marking and stem cell counts were reasonably close, reinforcing the accuracy of estimates for both stem cell numbers and unique integration sites.

When initiating the integration site analysis, one question was whether we would see in mice the dominance of an integration site similar to that found in HMGA2 in the human trial. In the human trial, integration of the Lentiglobin vector resulted in 3' end substitution of the gene as well as an increase in the rate of transcription initiation. The 3' end substitution removed a miRNA binding site in the 3' UTR. In mouse, HMGA2 has not been reported to be an oncogene. However, there are other examples insertional activation by 3' end substitution in the mouse. *Pim1* and *Gfi1* are oncogenes activated in T-cell lymphomas by substitution of their normal 3'UTRs, which contain miRNA binding sites [15, 43, 44]. Retroviral integration downstream of the oncogenes *Fgf3* (*int-2*) [45] and *c-myc* [46] has also been associated with oncogenic transformation. We thus might have seen enrichment of integration sites near these genes following transplantation into mice, but this was not the case. The closest site to *Pim1* was 80,427bp upstream of the 5' end in an intergenic region, the closest to *Fgf3* was over 3Mb upstream and lying within another gene, the closest to *c-myc* was over 10Mb upstream in an intergenic region. Of the seven most abundant sites recovered from the five mice, none were strong candidates for activation and clonal skewing by 3' end substitution of known growth control genes. Overall, the results in mice did not detectably recapitulate the clonal skewing associated with insertion in the HMGA2 third intron seen in the human β-thalassemia trial.

The main detectable difference in integration site distribution between pre-transplantation and post-transplantation samples was an effect on the frequency of

provirus accumulation in transcription units. We found that growth of transduced cells in mice or, to a greater extent, in methylcellulose was associated with a decrease in the favoring of integration in transcription units. We carried out a meta-analysis of previously published integration site datasets that have compared the proportion of lentiviral integration sites in transcription units pre-transplantation and post-transplantation into mice or humans. Shown in Table 4-4 are data from three published studies where cells were transduced by a lentiviral vector and transplanted into mice or humans, and vector integration sites were analyzed before and after transplantation. We found that two of the three studies (one to treat Wiskott-Aldrich syndrome in mice [47], and one to treat HIV in humans [27]) showed a decrease in the frequency of integration in transcription units after transplantation, though only one difference was statistically significant (Table 4-4). Evidence of selection against other retroviral elements integrated in genes can also be found. Reduced preference for integration in transcription units has been observed in patients chronically infected with HTLV, compared with *ex vivo* infections [48]. Similarly, it has been reported that evolutionarily older endogenous retroviruses are found less frequently in transcription units, in both mouse and human [49-53], likely reflecting selection against cells bearing integrated proviruses in transcription units. A potential explanation for our results would be that integration of lentiviral vectors into transcription units disrupted expression of the gene, and that this more often leads to a fitness cost for that cell than a fitness advantage. The effect could be at the transcriptional level, involving a change in mRNA levels, or via production of

Study	Reference	Vector	Promoter	Time of growth *in vivo*	Percentage in gene pre-transplantation	Percentage in gene post-transplantation	P value
This study		HIV SIN	Beta-globin	9 months	67.7	51.2	0.014
Lentiviral transduction of Lin- cells to treat a mouse model of WAS	[47]	HIV SIN	WASp	5 or 10 weeks	70.1	59.2	0.192
Lentiviral transduction of human CD4+ T cells to deliver an antisense HIV env gene	[27]	HIV full LTR	LTR	6, 14, 28 or 32 weeks	80.7	78.2	0.033
Lentiviral transduction of bone marrow and transplantation into tumor-prone mice	[13]	HIV SIN	PGK	6-10 months (mice sacrificed when tumors observed)	56.0	58.0	N/A

Table 4-4. Frequency of integration in transcription units in published studies.

abnormal proteins. Maruggi *et al.* [12] showed that lentiviral vectors caused transcriptional deregulation (both up- and down-regulation) of genes within 200kb of integration events. The frequency of this effect for lentiviral vectors was lower than for gammaretroviral, and the nature of the internal promoter appeared to be important. Thus effects on gene dosage may have resulted in some cells having a growth disadvantage that became evident after long-term proliferation.

In summary, our data show that gene transfer with the Lentiglobin vector used in the human clinical trial was not associated with clonal skewing during reconstitution of mice. However, our observations are consistent with a model in which the integration of a lentiviral vector in or near genes influenced the target cell via effects on gene activity. Surprisingly, however, it seems that the most common consequence in this study was a growth disadvantage, since cells with integrations within genes appear to have been selected against during growth. Following the findings in the human β-thalassemia trial, the results of this study are reassuring, but suggest more detailed analysis of the effects of integrating lentiviral vectors on cellular gene expression and associated fitness costs may be warranted.

4.6 Acknowledgements

We are grateful to members of the Bushman laboratory for help and suggestions. This work was supported by NIH grants AI052845 and AI082020. Keshet Ronen was supported by NIH training grant T32 AI-07324-17.

4.7 References

1. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, Gross F, Yvon E, et al. (2000) Gene therapy of human severe combined immunodeficiency (SCID)-X1

disease. Science 288(5466): 669-672.

2. Aiuti A, Slavin S, Aker M, Ficara F, Deola S, et al. (2002) Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. Science 296(5577): 2410-2413.

3. Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, et al. (2006) Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat Med 12(4): 401-409.

4. Cartier N, FAU - Hacein-Bey-Abina S, Hacein-Bey-Abina S, FAU - Bartholomae CC, Bartholomae CC, et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. - Science.2009 Nov 6;326(5954):818-23. (1095-9203 (Electronic)).

5. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, et al. (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science 302(5644): 415-419.

6. Hacein-Bey-Abina S, von Kalle C, Schmidt M, Le Deist F, Wulffraat N, et al. (2003) A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. N Engl J Med 348(3): 255-256.

7. Deichmann A, Hacein-Bey-Abina S, Schmidt M, Garrigue A, Brugman MH, et al. (2007) Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. J Clin Invest 117(8): 2225-2232.

8. Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, et al. (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest 118(9): 3132-3142.

9. Stein S, Ott MG, Schultze-Strasser S, Jauch A, Burwinkel B, et al. (2010) Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nat Med .

10. Naldini L, Blomer U, Gallay P, Ory D, Mulligan R, et al. (1996) In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science 272: 263-267.

11. Raineri I, Senn HP. (1992) HIV-1 promotor insertion revealed by selective detection of chimeric provirus-host gene transcripts. Nucleic Acids Res 20(23): 6261-6266.

12. Maruggi G, Porcellini S, Facchini G, Perna SK, Cattoglio C, et al. (2009) Transcriptional enhancers induce insertional gene deregulation independently from the vector type and design. Mol Ther 17(5): 851-856.

13. Montini E, Cesana D, Schmidt M, Sanvito F, Ponzoni M, et al. (2006) Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. Nat Biotechnol 24(6): 687-696. 10.1038/nbt1216.

14. Montini E, Cesana D, Schmidt M, Sanvito F, Bartholomae CC, et al. (2009) The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. J Clin Invest 119(4): 964-975.

15. Coffin JM, Hughes SH, Varmus HE. (1997) Retroviruses. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

16. De Palma M, Montini E, Santoni de Sio FR, Benedicenti F, Gentile A, et al. (2005) Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoetic cells. Blood 105: 2307-2315.

17. Wang GP, Garrigue A, Ciuffi A, Ronen K, Leipzig J, et al. (2008) DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. Nucleic Acids Res 36(9): e49.

18. Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C, Hubank M, et al. (2008) Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. J Clin Invest 118(9): 3143-3150.

19. Bushman FD. (2007) Retroviral integration and human gene therapy. J Clin Invest 117(8): 2083-2086.

20. Wu X, Li Y, Crise B, Burgess SM. (2003) Transcription start regions in the human genome are favored targets for MLV integration. Science 300(5626): 1749-1751.

21. Cattoglio C, Facchini G, Sartori D, Antonelli A, Miccio A, et al. (2007) Hot spots of retroviral integration in human CD34+ hematopoietic cells. Blood 110(6): 1770-1778.

22. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell

110(4): 521-529.

23. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. (2007) HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res 17(8): 1186-1194.

24. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, et al. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol 2(8): E234.

25. Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, et al. (2005) Genome-wide analysis of retroviral DNA integration. Nat Rev Microbiol 3(11): 848-858.

26. Levine BL, Humeau LM, Boyer J, MacGregor RR, Rebello T, et al. (2006) Gene transfer in humans using a conditionally replicating lentiviral vector. Proc Natl Acad Sci U S A 103(46): 17372-17377.

27. Wang GP, Levine BL, Binder GK, Berry CC, Malani N, et al. (2009) Analysis of lentiviral vector integration in HIV+ study subjects receiving autologous infusions of gene modified CD4+ T cells. Mol Ther .

28. Nishino J, Kim I, Chada K, Morrison SJ. (2008) Hmga2 promotes neural stem cell self-renewal in young but not old mice by reducing p16Ink4a and p19Arf expression. Cell 135(2): 227-239.

29. Lee YS, Dutta A. (2007) The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. Genes Dev 21(9): 1025-1030.

30. Mayr C, Hemann MT, Bartel DP. (2007) Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. Science 315(5818): 1576-1579.

31. Cleynen I, Van de Ven WJ. (2008) The HMGA proteins: A myriad of functions (review). Int J Oncol 32(2): 289-305.

32. Baldassarre G, Fedele M, Battista S, Vecchione A, Klein-Szanto AJ, et al. (2001) Onset of natural killer cell lymphomas in transgenic mice carrying a truncated HMGI-C gene by the chronic stimulation of the IL-2 and IL-15 pathway. Proc Natl Acad Sci U S A 98(14): 7970-7975.

33. Wu X, Burgess SM. (2004) Integration target site selection for retroviruses and transposable elements. Cell Mol Life Sci 61(19-20): 2588-96.

34. Ciuffi A, Ronen K, Brady T, Malani N, Wang G, et al. (2009) Methods for integration site distribution analyses in animal cell genomes. Methods 47(4):

261-268.

35. Berry C, Hannenhalli S, Leipzig J, Bushman FD. (2006) Selection of target sites for mobile DNA integration in the human genome. PLoS Comput Biol 2(11): e157.

36. Chao A. (2001) An overview of closed capture-recapture models. Journal of Agricultural, Biological, and Environmental Statistics 6(2): 158-175.

37. Akagi K, Suzuki T, Stephens RM, Jenkins NA, Copeland NG. (2004) RTCGD: Retroviral tagged cancer gene database. Nucleic Acids Res 32(Database issue): D523-7.

38. Tarbe NG, Rio MC, Hummel S, Weidle UH, Zoller M. (2005) Overexpression of the small transmembrane and glycosylated protein SMAGP supports metastasis formation of a rat pancreatic adenocarcinoma line. Int J Cancer 117(6): 913-922.

39. Tarbe NG, Rio MC, Weidle UH. (2004) SMAGP, a new small trans-membrane glycoprotein altered in cancer. Oncogene 23(19): 3395-3403.

40. Miller SC. (2002) Hematopoietic reconstitution of irradiated, stem cell-injected mice: Early dynamics of restoration of the cell lineages of the spleen and bone marrow. J Hematother Stem Cell Res 11(6): 965-970.

41. Harrison DE, Zhong RK. (1992) The same exhaustible multilineage precursor produces both myeloid and lymphoid cells as early as 3-4 weeks after marrow transplantation. Proc Natl Acad Sci U S A 89(21): 10134-10138.

42. Harrison DE, Lerner CP. (1991) Most primitive hematopoietic stem cells are stimulated to cycle rapidly after treatment with 5-fluorouracil. Blood 78(5): 1237-1240.

43. Selten G, Cuypers HT, Berns A. (1985) Proviral activation of the putative oncogene pim-1 in MuLV induced T-cell lymphomas. EMBO J 4(7): 1793-1798.

44. Dabrowska MJ, Dybkaer K, Johnsen HE, Wang B, Wabl M, et al. (2009) Loss of MicroRNA targets in the 3' untranslated region as a mechanism of retroviral insertional activation of growth factor independence 1. J Virol 83(16): 8051-8061.

45. Dickson C, Smith R, Brookes S, Peters G. (1990) Proviral insertions within the int-2 gene can generate multiple anomalous transcripts but leave the protein-coding domain intact. J Virol 64(2): 784-793.

46. Corcoran LM, Adams JM, Dunn AR, Cory S. (1984) Murine T lymphomas in which the cellular myc oncogene has been activated by retroviral insertion. Cell 37(1): 113-122.

47. Mantovani J, Charrier S, Eckenberg R, Saurin W, Danos O, et al. (2009) Diverse genomic integration of a lentiviral vector developed for the treatment of wiskott-aldrich syndrome. J Gene Med 11(8): 645-654.

48. Meekings KN, Leipzig J, Bushman FD, Taylor GP, Bangham CR. (2008) HTLV-1 integration into transcriptionally active genomic regions is associated with proviral expression and with HAM/TSP. PLoS Pathog 4(3): e1000027.

49. Smit AF. (1999) Interspersed repeats and other momentos of transposable elements in mammalian genomes. Curr Opin Genet Dev 9(6): 657-63.

50. Lander E. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921.

51. Medstrand P, van de Lagematt LN, Mager DL. (2002) Retroelement distributions in the human genome: Variations associate with age and proximity to genes. Genome Res. 12: 1483-1495.

52. Zhang Y, Maksakova IA, Gagnier L, van de Lagemaat LN, Mager DL. (2008) Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements. PLoS Genet 4(2): e1000007.

53. Brady T, Lee YN, Ronen K, Malani N, Berry CC, et al. (2009) Integration target site selection by a resurrected human endogenous retrovirus. Genes Dev 23(5): 633-642.

CHAPTER 5 – NOVEL HOST FACTORS IN HIV INTEGRATION SITE SELECTION

5.1 Abstract

Three genome-wide siRNA screens have recently been published identifying host factors necessary for HIV infection. There has been much interest in characterizing in more detail the effects of identified factors and possible mechanisms of action. We analyzed a number of these host factors, and additional candidate host factors identified by other means, that appear to act at the integration steps of the replication cycle. In previous studies, LEDGF/p75 was found to be important for efficient infection, and to act as a targeting factor directing HIV integration to active transcription units. We used 454 sequencing of integration sites to ask whether any of the newly identified factors that are important for efficient integration also influence integration targeting. We carried out bioinformatic analysis of HIV integration sites isolated from cells treated with siRNAs against PRPF38A, MAP4 and SETD2. These data suggest integration site selection is altered when under these conditions, though whether these effects are due to depletion of these factors or a more indirect mechanism is unclear. Common themes emerge from the analysis of these knockdowns, with diminished enrichment of integration sites in transcription units and gene-rich regions compared with integration in control cells. Surprisingly, the patterns differ from that observed in integration sites from LEDGF/p75-depleted

cells, suggesting that LEDGF/p75 may not be the only host factor responsible for the characteristic genomic distribution of HIV integration sites.

5.2 Introduction

Integration is a necessary step of the HIV lifecycle, whereby the reverse-transcribed viral genome becomes covalently joined to the genome of the infected cell. In vitro, the viral protein integrase (IN) is sufficient to mediate the integration of donor into target DNA [1-3], and the reaction shows weak specificity for target sequence [4-6]. However, in the cell, IN catalysis takes place in the context of a large nucleoprotein complex, the pre-integration complex (PIC), into host chromatin in the nuclear environment. A number of host proteins have been shown to interact with the PIC [7-11] with potential roles in infection. Furthermore, integration occurs in a non-random distribution in the host genome, which is not accounted for by IN's weak sequence preference [12-14]. Integration site preferences vary between retroviral classes, with lentiviruses such as HIV showing a preference for active transcription units and gammaretroviruses such as MLV showing a preference for CpG islands and gene 5' ends [13, 14]. The fact that these preferences track with retroviral genus but are not seen in *in vitro* integration by purified IN or PICs suggests that viral components of the PIC likely interact with host proteins or chromatin structure to influence the distribution of integration events in the genome.

Indeed, this prediction has been confirmed by studies with the host protein LEDGF/p75 [15, 16]. LEDGF/p75 was identified as a binding partner of lentiviral IN [11, 17], and was shown to mediate IN chromatin binding [18, 19], protect IN from

proteasomal degradation [20] and stimulate IN *in vitro* activity [11]. Human and mouse models of stringent LEDGF/p75 depletion have shown that LEDGF/p75 is a necessary integration factor: in a knockdown or knockout integration by HIV or a related lentivirus EIAV were blocked 5- to 30-fold [21-23]. Furthermore, integration in the absence of LEDGF/p75 showed a different distribution around the genome from that in wild-type cells. In LEDGF/p75-depleted cells, the preference for integration in transcription units was diminished [22-24]. Additionally, the GC content of the sequence around integration sites was observed to increase and integration in CpG islands became favored in the absence of LEDGF/p75. This has led to a tethering model of integration placement, whereby a chromatin-bound host factor interacts with a viral component of the PIC and tethers integration events to the regions around its binding sites. In the case of LEDGF/p75, this model has been further confirmed through the demonstration of integration retargeting by expression of fusion proteins containing the IN-binding portion of LEDGF/p75 fused to alternative DNA binding domains (as described in Chapter 3) [25, 26].

However, studies of LEDGF/p75 in integration targeting leave some open questions. Firstly, LEDGF/p75's cellular role, binding partners and mechanism of chromatin interaction remain partially understood, raising the question of whether other factors might participate in this tethering mechanism. Secondly, while depletion of LEDGF/p75 has marked and significant effects on the distribution of lentiviral integration sites, sites are not randomly distributed in its absence, suggesting other factors may contribute.

We sought to investigate alternative integration site determinants. In considering candidate factors we reasoned that novel tethering factors would likely phenocopy the known tethering factor, LEDGF/p75. We therefore focused initially on host proteins whose depletion from the cell, like LEDGF/p75, led to an infection block at the step of integration, and analysed the distribution of HIV vector integration sites in cells depleted for these factors. A large number of candidate factors was provided in the form of hits from a recent genome-wide siRNA screen by König *et al.* that identified factors necessary for HIV infection [27]. We report here integration site data from cells depleted by siRNA knockdown for two factors identified as hits by König *et al*. PRPF38A is a splicing factor, annotated in the human genome by its homology to the yeast protein PRP38, which is required for spliceosome maturation [28]. It was identified as a factor necessary for integration in the König *et al*. screen. MAP4 is a microtubule-associated protein involved in microtubule assembly and cell cycle progression [29, 30]. It was independently identified in two genome-wide siRNA screens [27, 31] and has also been reported to interact with LEDGF/p75 by yeast-two-hybrid (Sumit Chanda, personal communication). We additionally investigated SETD2/HYPB, a histone methyl-transferase that methylated histone 3 on lysine 36 [32, 33], a modification that is found in the bodies of transcription units and closely follows the distribution of HIV integration sites in the genome [26]. SETD2 recruits to active genes several other proteins involved in mRNA processing and export [34]. One of these is IWS1, which has been shown to bind LEDGF/p75 (Katherine Jones, personal communication).

Knockdown of SETD2 led to a reduction in HIV infection efficiency, though the gene had not been identified as a hit in genome-wide siRNA screens. We found that depletion of PRPF38A, SETD2 and MAP4 resulted in an altered distribution of integration sites, and common themes emerged from the changes: knockdown of all genes led to a slight reduction in the frequency of integration in transcription units and a shift of integration to less gene-dense and GC-rich regions. In data not shown, from Troy Brady in the lab, knockdown of nuclear import factors TNPO3 and ANAPC2 showed similar effects. In order to rule out off-target effects, we expressed an siRNA-resistant form of SETD2, and found that the infection block induced by siRNA-treatment was not rescued. Thus, the significance of these effects is unclear. Possible interpretations are discussed.

5.3 Materials and Methods

Cell culture and transfection

293T cells were cultured in Dulbecco's Modified Eagle Medium with Glutamax (Invitrogen), 10% FBS (Sigma) and 50μg/ml gentamycin (Sigma). For siRNA treatment, cells were reverse-transfected. Cells were seeded in 12-well plate format, 100,000 cells to a well in antibiotic-free DMEM with 10% FBS. siRNAs were incubated with Optimem serum-free medium (Invitrogen) and RNAiMax transfection reagent (Invitrogen), as per manufacturer's instructions, and added to the cells at the time of seeding. 37.6pmol siRNA and 1.88ul RNAiMax were used per well. For co-transfection of siRNAs and rescue plasmids, reverse transfection was

carried out in the same way, using 37.6pmol siRNA, 500ng plasmid DNA and 1.88ul RNAiMax were used per well.

siRNAs used

siRNAs against SETD2 were from Qiagen (siRNA 1 in the text is HYPB_1, catalog number SI00103292) and Ambion (siRNA 2 in the text has sequence GUGAAGGAGUAUGCACGAAtt). siRNAs against PRPF38A and MAP4 were from Qiagen (catalog numbers SI00395808, SI00395815, SI00627809, SI00627816). The control luciferase siRNA was from Qiagen (GL2, catalog number 1022070).

siRNA-resistant cDNA cloning

SETD2 cDNA was purchased from Open Biosystems (clone ID 40125715). Site-directed mutagenesis was performed using the Quikchange kit (Stratagene). Two sets of 7 synonymous mutations were introduced separately, to each of the regions targeted by the siRNAs used. The coding sequence was then amplified to incorporate restriction enzyme sites (primers kr225, AGTCCAagatctagaGAAAGAAGAGGCAAGTATTCTTC and kr227, agtccaGTCGACctcgagTCACTCTAATTCAGTGTCCTCTTTGG). The amplicon was digested with BglII and SalI and inserted into a plasmid containing a 3xFlag tag, digested with BamHI and SalI. Flag-tagged SETD2 was then cut out using AgeI and SalI and ligated into a vector based on the MLV-based CMV-promoter-driven expression plasmid pLNCX (kind gift of Paul Bates), previously engineered to carry a WPRE.

Viral particle production and infection

VSV-G pseudotyped HIV vector particles were produced by Lipofectamine transfection of 293T cells with p156RRLsin-PPTCMVGFPWPRE [35], the packaging construct pCMVdeltaR9 [36], and the vesicular stomatitis virus G-producing pMD.G construct. Viral supernatant was harvested 38 hours after transfection, filtered through 0.22µm filters, concentrated by filtration through a Centricon, treated with DNase I, and stored frozen at -80°C. HIV titer was quantified by p24 ELISA.

Cells were infected 48h after transfection: medium from the transfection was removed and replaced with DMEM + 10% FBS and 50µg/ml gentamycin containing 50-100ng p24 VSV-G pseudotyped virus. Infection mix was left on the cells overnight and then replaced with fresh medium. Infection was allowed to proceed for 48h, after which cells were trypsinized and harvested for FACS, proviral Q-PCR, gene expression analysis by Q-RT-PCR, or integration site analysis.

Q-PCR

For quantification of integrated proviruses, a two-step Alu-PCR assay was used, described in reference [37]. This consists of an intitial round of amplification from genomic Alu repeats to viral *Gag* sequence, followed by Q-PCR specific for an amplicon in R-U5 with molecular beacon probes on an Applied Biosystems 7500 Fast Realtime PCR instrument. Samples were run in triplicate.

For gene expression analysis, total RNA was extracted from cells using the Illustra RNAspin kit (GE healthcare). 50ng RNA from each sample was reverse transcribed using the High Capacity RNA to cDNA kit (Applied Biosystems). cDNA

was diluted 100-fold and QPCR performed in triplicate on each sample with commercial primers and Taqman MGB FAM-labeled probes (Applied Biosystems, SETD2 assay ID: Hs00383438_m1) and Taqman Fast Universal Mastermix (Applied Biosystems, catalog number 4352042). Q-PCR was also performed on all samples with GAPDH primers and probe (Applied Biosystems catalog number: 402869). SETD2 expression was calculated by the ΔΔCt method. An Applied Biosystems 7500 Fast Realtime PCR instrument was used.

Western blot

Cell pellets were lysed in 1X RIPA buffer with protease inhibitors (Roche, catalog number 11697498001), lysates mixed with SDS buffer and subjected to SDS polyacrylamide gel electrophoresis. Flag was detected using anti-Flag M2-peroxidase conjugated (Sigma, catalog number A8592). β-tubulin was used as a loading control (Abcam ab21058). SETD2 detection with Abcam antibody ab69836 was attempted but gave only non-specific bands. Visualization was by chemiluminesence, using Supersignal West Chemiluminescent substrate (Pierce, catalog number PI34080).

Integration site analysis

Integration sites were isolated and sequenced by ligation-mediated PCR essentially as described previously [38]. Genomic DNA was extracted from infected cells using the Qiagen DNeasy tissue extraction kit. Up to 2µg of DNA from each infection was digested overnight using MseI. This was followed by digestion to prevent amplification of internal viral fragments (from the 5' LTR) and plasmid backbone with SacI and DpnI. Linkers were then ligated onto digested products

(oligonucleotide sequences listed below) and nested PCR performed from ligation products. Nested PCR primers contained 8 nt barcode sequences between the sequencing primer and LTR-binding portions. These enabled pooling of all PCR products into one sequencing reaction and subsequent separation of sequences by decoding the barcodes. Samples were sequenced on the Roche 454 GS-FLX instrument at the University of Pennsylvania.

Integration sites were judged to be authentic if the sequences had a best unique hit when aligned to the murine (mm8 draft) using BLAT, and the alignment began within 3bp of the viral LTR end and had >98% sequence identity. Detailed statistical methods are described in [6].

To control for possible biases in isolating integration sites due to restriction enzyme sequence distribution, three or ten matched random controls were computationally generated for each experimental integration site that were the same distance from the closest MseI restriction site as the experimental site.

Integration site counts in various genomic annotations were compared with matched random controls by the Fisher's exact test. Additionally, multiple regression models for integration intensity were applied, as described in [6]. Analysis was carried out in the R statistical package (http://www.r-project.org).

5.4 Results

HIV infection is inhibited by knockdown of MAP4, PRPF38A and SETD2

König and colleagues recently conducted a genome-wide siRNA screen to identify human genes necessary for HIV infection [27]. 293T cells were treated with

siRNAs against a panel of around 20,000 genes and hits identified as genes that, when knocked down by two or more independent siRNAs, led to a block in HIV reporter gene expression. Hits were prioritized for downstream analysis based on several lines of evidence such as cellular toxicity of knockdown, co-expression with CD4, CXCR4 and CCR5, and evidence of interaction with HIV proteins. For genes passing these filters, the infection block induced by siRNA treatment was placed in a stage of the viral lifecycle. We focused on hits from the screen whose effects had been mapped to the integration step of the lifecycle. MAP4, a microtubule associated protein, and PRPF38A, a splicing factor, were two such factors. Additionally, we investigated the potential role of SETD2/HYPB, a histone methyltransferase suggested to interact with a LEDGF/p75-binding protein.

293T cells were treated with siRNAs directed against PRPF38A, MAP4, SETD2 or luciferase as a control, and 48h later infected with VSV-G pseudotyped GFP-expressing HIV. Cells were harvested 48h later for FACS analysis to measure infection efficiency, and for integration site analysis. Table 5-1 summarizes the genes studied and their effects on HIV infection. In agreement with the results of König *et al*., knockdown of both MAP4 and PRPF38A caused a decrease in susceptibility to HIV infection (on average 1.67-fold and 1.35-fold respectively). Knockdown of SETD2 with two different siRNAs also led to a decrease in HIV infection efficiency (on average a 1.51-fold decrease in GFP expression).

<u>HIV integration site selection is modified by knockdown of PRPF38A, SETD2 and MAP4</u>

Gene	Information	Known functions	siRNA used	Fold Reduction HIV Infection (FACS)
Control (luciferase)			GL2	1.00
PRPF38A	Identified in siRNA screen (integration factor)	Spliceosome	PRPF38A_2	1.67
			PRPF38A_3	nd
MAP4	Identified in siRNA screen (integration factor); binds LEDGF/p75 by Y2H.	Microtubule binding	MAP4_3	1.20
			MAP4_4	1.49
SETD2	Binding partner of IWS1, which interacts with LEDGF/p75	H3K36 methyltransferase	SETD2_1	1.61
			SETD2_2	1.40

Table 5-1. Effects of genes studied on HIV infection efficiency. 293T cells were transfected with the siRNAs shown 48h prior to infection with VSV-G pseudotyped GFP-expressing HIV. The percentage of cells expressing GFP was determined by FACS 48h post-infection. Numbers are expressed as fold reduction of this value compared with control-infected cells.

Having confirmed that knockdown of MAP4, PRPF38A and SETD2 reduced the efficiency of HIV infection, we examined the effect of these potential cofactors on integration site selection. Integration site amplification was carried out essentially as described previously [38], sites aligned to the human genome and nearby genomic features annotated. The number of integration sites analyzed from each siRNA treatment and summaries of genomic features near these sites are shown in Table 5-2. For reference, a dataset from 293Ts stably knocked down for LEDGF/p75 (the siLL cells from [24]) is shown, as well as a set of computationally generated random controls (see methods). One of the goals of this study was to identify factors that may participate in LEDGF/p75 tethering. Depletion of factors that cooperate with LEDGF/p75 would be expected to alter integration targeting in the same way as LEDGF/p75 depletion does. We therefore started by examining the preference for integration in genes. Table 5-2 summarizes the proportions of integration sites falling in transcription units from cells treated with various siRNAs. Consistent with previous data, HIV integration from control cells was favored in the bodies of transcription units, with 72.1% of sites falling in RefSeq genes.[12], and this preference was reduced in LEDGF/p75 knockdown cells [24]. Integration sites from cells treated with siRNAs against MAP4, PRPF38A and SETD2 all showed slight decreases in the frequency of integration in transcription units, with MAP4 knockdown showing the weakest effect (67.6% and 65.7% in genes depending on the siRNA used), PRPF38A intermediate (61.8% and 62.7% in genes) and SETD2 the strongest effect (59.5% in genes). Comparing these proportions to those in control

Data set	Integration sites	% in genes	Average GC % in 5kb	Average number genes in 1Mb
GL2	4152	72.1	40.5	20.2
PRPF38A_2	1591	61.8***	38.0***	13.8***
PRPF38A_3	721	62.7**	38.7***	15.1***
MAP4_3	3730	67.6	39.4***	17.5***
MAP4_4	3418	65.7**	39.9***	17.9***
SETD2_1	2271	59.5***	39.1***	17.0***
LEDGF_siLL	468	59.2***	41.8***	18.9
Random	12954	35.6***	39.7***	10.1***

Table 5-2. Effect of cofactor knockdown on the genomic distribution of HIV integration sites. Cells were transfected with the siRNAs shown and 48h later infected with VSV-G pseudotyped HIV. Integration sites were isolated 48h post-infection, aligned onto the genome and annotated with respect to the genomic features shown. Asterisks denote statistical significance as determined by a c-logit test applied to a logistic regression model, or Mann-Whitney U test as appropriate. * $p<0.05$, ** $p<0.01$, ***$p<0.0001$.

cells achieved statistical significance, but as shown in Table 5-2, the changes were of smaller magnitude than those seen in LEDGF/p75-depleted cells. It should also be noted that the LEDGF/p75 knockdown shown here is a partial knockdown – more stringently depleted models have been made in other cell types that show a greater reduction in the frequency of integration in genes (see Chapter 2).

Another effect observed in LEDGF/p75-depleted cells is a shift of integration sites into regions of higher GC content [22-24]. We therefore wondered whether knockdown of these factors might show the same effect, supporting the idea that these factors are part of the LEDGF/p75 machinery. Table 5-2 shows the average GC content of 5kb regions surrounding integration sites isolated from control and siRNA-treated cells. It can be seen that the GC content of regions surrounding integration sites from different treatments varied, with LEDGF/p75 knockdown showing an increase in GC content as previously reported. However all other knockdowns showed decreases in the average GC content. While these differences did achieve statistical significance when compared to control cells, the magnitude of the changes was very small (ranging from 41.8% GC in LEDGF/p75 knockdown cells to 38.0% in PRPF38A knockdown cells). A more sensitive way to analyze this type of data than to examine the average value surrounding each site is to compare the distributions of values around sites from two conditions. Since siRNAs targeted against the same gene showed similar effects, we pooled integration sites from different siRNAs to generate such graphs. Figure 5-1 shows this analysis. The GC content in the 5kb surrounding each integration site was calculated, then sites from control and treated

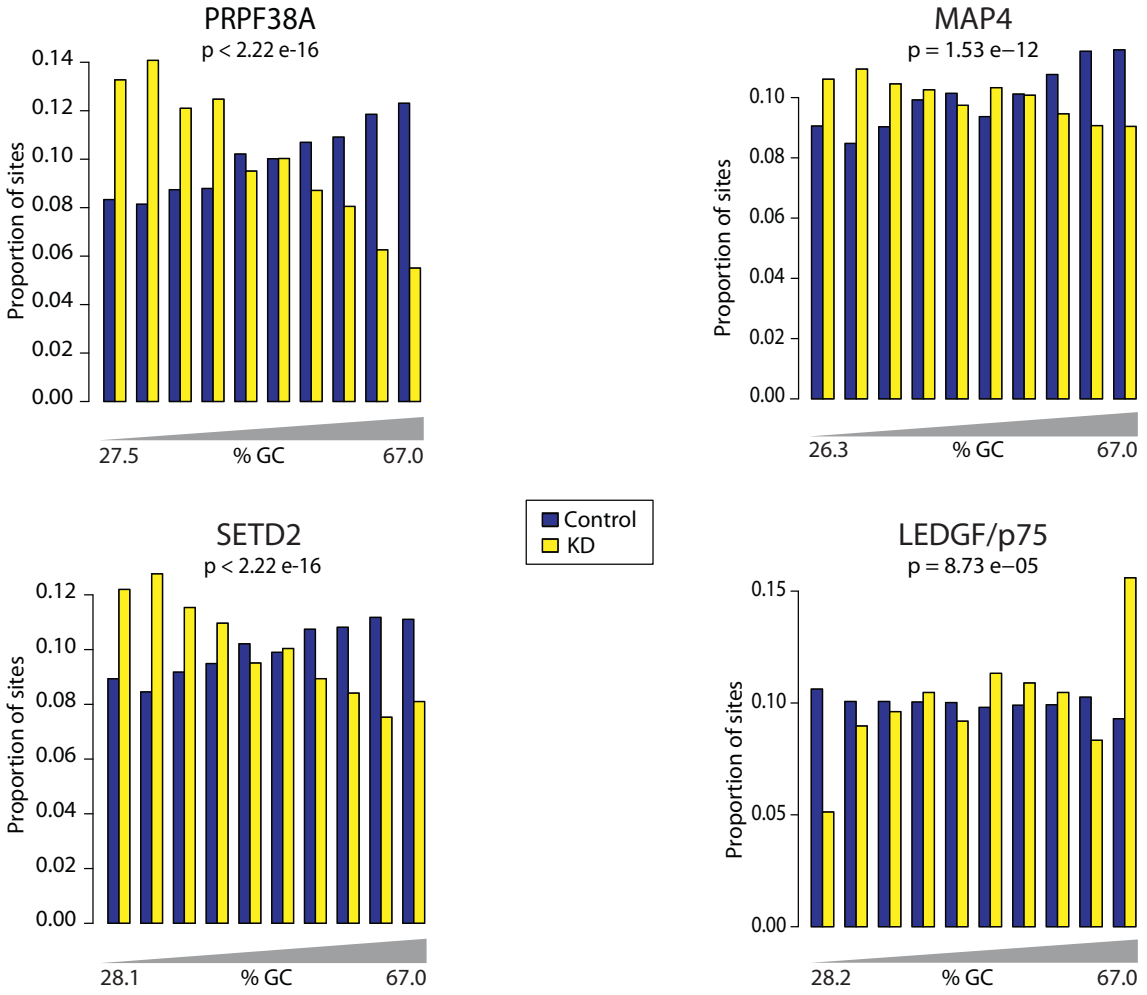

Figure 5-1. The effect of cofactor knockdown on integration with respect to GC content. HIV vector integration sites from cells treated with siRNAs against each of the genes shown were isolated and mapped onto the genome. The GC content in a 5kb window around each site was calculated. Sites from control and knockdown sets were combined and split into ten bins of increasing GC content with equal numbers of sites in each bin. The proportion of each set found in each bin is shown. P values were determined using the likelihood ratio statistic for the logistic regression model. LEDGF/p75 knockdown data are from stable knockdown 293T cells as in [19].

cells combined and broken into ten bins of increasing GC content. The proportion of sites from each treatment found in each bin is displayed. It can be seen that, whereas LEDGF/p75 depletion led to an increase in the GC content around sites, knockdown of PRPF38A, SETD2 and, to a lesser extent, MAP4, had the opposite effect. Though these changes are all slight, they are statistically significant, and the difference between LEDGF/p75 knockdown and the other knockdowns argues against the idea that knockdown of the genes studied here disrupts the same complex as LEDGF/p75 depletion.

A number of genomic features tend to correlate in the genome, for example regions with a high GC content tend to also be gene-rich. We therefore analyzed the gene density of 1Mb windows surrounding integration sites from cells treated with different siRNAs. The results are summarized in Table 5-2 and shown graphically in Figure 5-2, analyzed in the same way as GC content in Figure 5-1. In agreement with the effect on GC content, we observed that knockdown of PRPF38A, SETD2 and MAP4 led to integration sites lying in less gene-dense regions. LEDGF/p75 knockdown, however, had no effect on gene density surrounding integration sites, again arguing against the idea that any of the novel factors tested here participate in LEDGF/p75-mediated integration tethering.

The effects of SETD2 knockdown may be off-target

siRNA treatment is known to cause off-target effects, such as knockdown of additional genes with sequence similarity to the target gene, or induction of an interferon response [39]. Using two siRNAs targeted against each gene increased our

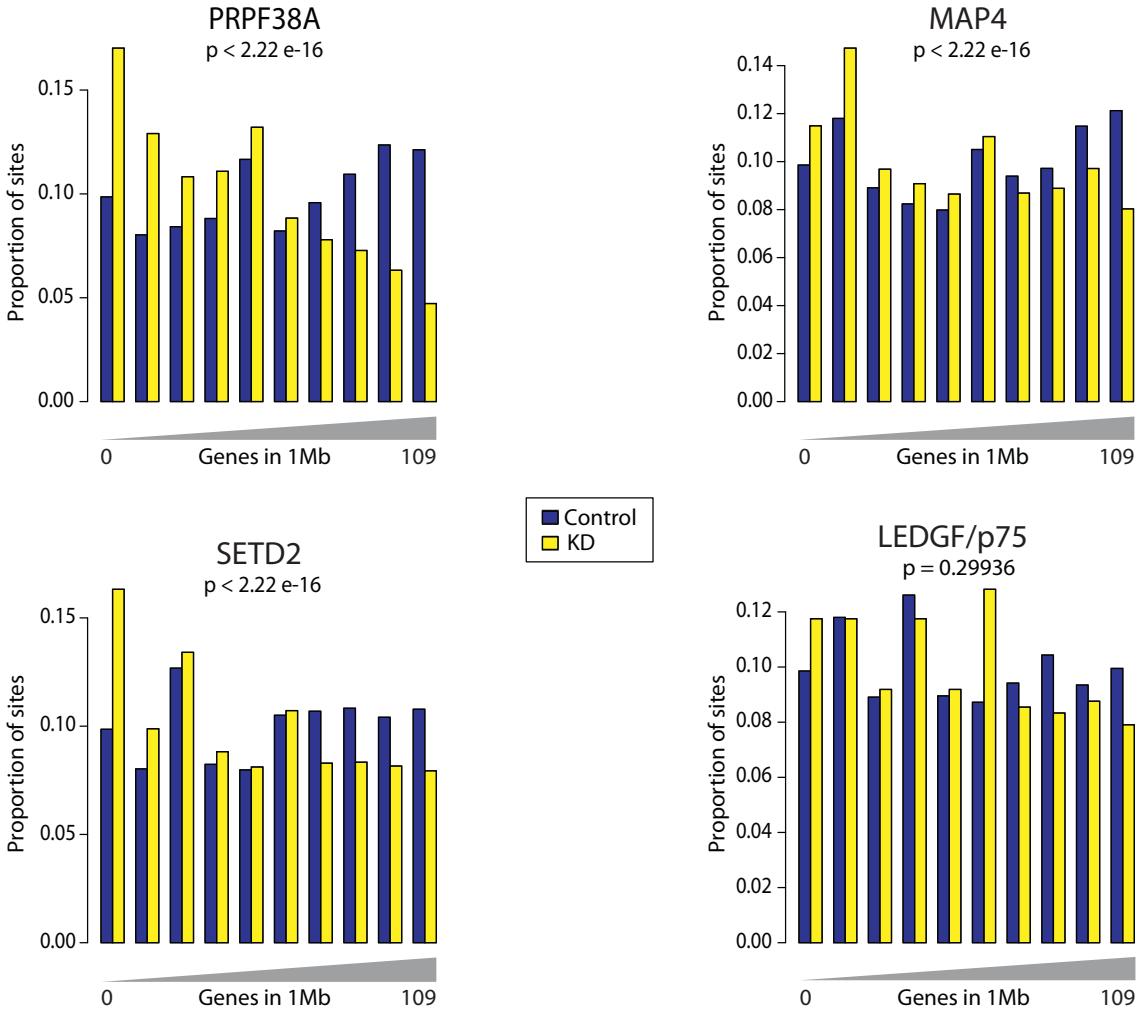

Figure 5-2. The effect of cofactor knockdown on integration with respect to gene density. HIV vector integration sites from cells treated with siRNAs against each of the genes shown were isolated and mapped onto the genome. The number of RefSeq genes in a 1Mb window around each site was calculated. Sites from control and knockdown sets were combined and split into ten bins of increasing gene density with equal numbers of sites in each bin. The proportion of each set found in each bin is shown. P values were determined using the likelihood ratio statistic for the logistic regression model. LEDGF/p75 knockdown data are from stable knockdown 293T cells as in [19].

confidence that the effects on infection and integration site selection were authentically due to depletion of the proteins of interest. However, we sought to confirm the specificity of the effect by delivering to knockdown cells an siRNA-resistant clone of the gene of interest, and testing for the block to infection. This allows expression of the protein to be rescued, but any potential off-target effects induced by the siRNA treatment to remain.

cDNAs of SETD2 were engineered to contain 7 synonymous mutations in the 21bp recognition sequences of the two siRNAs that inhibited HIV infection. These mutated cDNAs were flag-tagged for ease of detection and cloned into a mammalian expression vector containing a CMV promoter. 293T cells were cotransfected with control or SETD2 siRNAs and the rescue constructs, or empty vector as a control. 48h later, cells were infected as before, and 48h after infection cells were harvested. SETD2 transcription was measured by quantitative RT-PCR and expression of the flag-tagged rescue constructs was measured by Western blot. Figure 5-3A and B shows that the rescue constructs were successfully expressed at the level of RNA and protein. The susceptibility of cells to HIV infection was determined by measuring integrated viral copies by quantitative PCR. These results are displayed in Figure 5-3C. Cells transfected with the control siRNA and rescue constructs showed slight increases in infection efficiency compared with control cells, indicating that overexpression of SETD2 was not toxic or inhibitory to infection. Cells treated with the two siRNAs directed against SETD2 showed a roughly 3-fold reduction in the number of integrated proviruses. Cotransfection of rescue constructs did not lift this

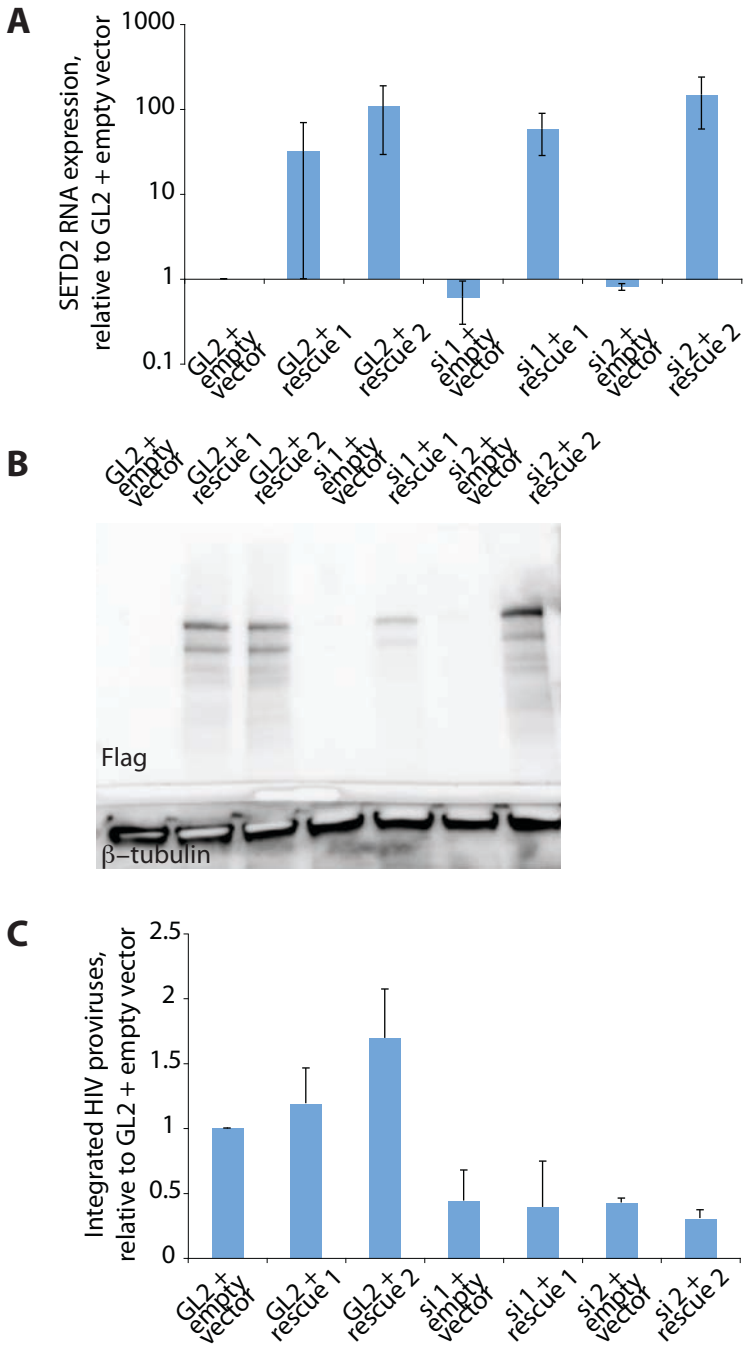

Figure 5-3. The infection block induced by SETD2 siRNA treatment is not rescued by overexpression of siRNA-resistant SETD2 cDNA. A. Expression of SETD2 RNA, measured by Q-RT-PCR. Values shown are means of two biological replicates, plus and minus standard deviation. **B.** Representative Western blot for Flag-tagged construct expression. **C.** Integrated HIV proviruses, measured by Q-PCR. Values shown are means of two biological replicates, plus standard deviation.

infection block. It remains possible that the Flag-tagged constructs used here were misfolded, Myc- and Flag-tagged SETD2 or portions of SETD2 have been used with apparent success in previous publications [40]. This result therefore fails to rule out the possibility that off-target effects are responsible for the inhibition of HIV infection by SETD2 knockdown.

Knockdown of other genes involved in HIV integration and nuclear import shows similar results

Experiments conducted by Troy Brady and Karen Ocwieja in the lab (data not shown) have examined the effect of other knockdowns on HIV integration site selection. Two more hits were selected from genes identified in the König siRNA screen. ANAPC2 is another gene identified in the screen as a factor that when knocked down reduced the level of HIV integration. It is a component of the anaphase promoting complex involved in cell cycle regulation [41]. TNPO3 (transportin SR2) is a member of the karyopherin β family of proteins that shuttles between the nucleus and cytoplasm and is involved in nuclear import of proteins such as SR splicing factors [42]. Knockdown of TNPO3 was shown to inhibit HIV infection at the stage of nuclear import in two recent genome-wide siRNA screens [27, 31]. It was also independently identified as interacting with HIV integrase [10] and confirmed to be required for nuclear import, though subsequent studies of its role in infection have suggested the involvement of the capsid protein [43].

Knockdown of ANAPC2 and TNPO3 in 293T cells, infection and integration site analysis was conducted as described above for SETD2, PRPF38A and MAP4.

The effects of knockdown of ANAPC2 and TNPO3 were found to be similar to those reported here. ANAPC2 and TNPO3 knockdown did not show significant effects on the frequency of integration within transcription units, but did lead to a reduction in the GC content and gene density of chromatin surrounding HIV integration sites.

Add-back experiments, as described here for SETD2, were completed for both TNPO3 and ANAPC2. In those cases, rescuing gene expression did mediate at least a partial rescue of the infectivity and integration site effect.

5.5 Discussion

Much attention has recently been paid to host factors necessary for HIV infection. Studying interactions between host and viral proteins is valuable in furthering our understanding of the HIV lifecycle and identifying potential therapeutic targets. One aspect of the lifecycle where host factors are known to be important is integration and the targeting of certain genomic features by integration events: LEDGF/p75 is required for efficient integration and determining much of the preference of lentiviruses for integration in transcription units. However, other factors may also be important, either as part of the LEDGF/p75 machinery, or as a parallel system still active in its absence.

In this study we investigated whether host proteins whose knockdown inhibits HIV infection at the step of integration also have an impact on integration site selection. We tested this by treating cells with siRNAs directed against three factors, PRPF38A, MAP4 and SETD2 whose knockdown reduced HIV infection efficiency. Small but statistically significant changes in the genomic distribution of integration

sites were observed, with broadly similar effects resulting from the three knockdowns: the frequency of integration in genes was slightly reduced, the GC content of 5kb windows around integration sites was reduced, and the gene density of 1Mb windows around integration sites was reduced.

The significance of these findings remains unclear. Importantly, the changes induced by knockdown of these proteins differ from the changes induced by LEDGF/p75 depletion. This argues against the idea that the factors tested were LEDGF/p75 cofactors. A number of interpretations could therefore be imagined. The first is that the proteins studied in this work all assist in the nuclear trafficking of the pre-integration complex, enabling nuclear import (eg. TNPO3) or integration (eg. PRPF38A, MAP4, SETD2, ANAPC2). This is known to be true for TNPO3, and suggested, though not demonstrated, for MAP4 and SETD2, by their purported interaction with LEDGF/p75 or its binding partners. Disruption of this trafficking might alter the placement of integration events in chromatin. Chromatin is known to be organized into higher order structures in the nucleus, with characteristic distributions of particular genomic regions [44]. Perhaps disruption of nuclear trafficking pathways, by depletion of any of the above factors, misdirects PICs into less gene-dense, GC-rich regions of the genome.

However, questions remain about off-target effects. The infection block induced by SETD2 siRNAs could not be rescued by siRNA-resistant cDNA expression, and the integration site effects observed in this system were very similar to those observed in others. Another possible interpretation of the data therefore

remains: that siRNA treatment resulted in knockdown of some unintended factor or induced an innate immune or stress response. In this situation, some other indirectly modulated factor could be the true mediator of integration targeting in gene-dense regions. Relatedly, such off-target effects could alter cellular chromatin structure to make gene-dense regions less accessible. The induction of an interferon response could be studied by assaying the expression of interferon-inducible genes following siRNA treatment. Similarly, cells could be treated with interferon-β prior to infection to see if the integration site selection was redistributed as reported here.

A related confounding effect could be that on-target depletion of these proteins, though not affecting HIV infection directly, results in global gene expression changes that then alter integration targeting, by modulating some tethering factor or causing changes in chromatin accessibility.

It should be noted that similar effects to those reported here have been observed when comparing integration site selection in other settings. Comparing dividing and growth-arrested IMR90 fibroblasts, Ciuffi *et al.* [45] found that HIV integration sites from dividing cells were in less gene-dense regions than growth arrested cells. This result could lend support to the idea that passage of the PIC through the nuclear pore, which would be necessary in non-cycling cells, may have an impact on the distribution of sites in chromatin. However, the opposite effect was reported by Brady *et al.*, who showed that integration sites in activated T-cells were in more gene-dense and GC-rich regions than integration sites in resting T-cells, which divide less [46]. Furthermore, Chapter 6 of this dissertation presents data that

pharmacological inhibition of integrase strand transfer also leads to similar effects on HIV integration site selection.

In all, these data suggest that HIV integration targeting may be shaped by variables in addition to LEDGF/p75 expression. Whether the proteins described here participate in this process, or their knockdown affected integration in some other way requires further study.

5.6 References

1. Panganiban AT, Temin HM. (1984) The retrovirus pol gene encodes a product required for DNA integration: Identification of a retrovirus int locus. Proc. Natl. Acad. Sci. USA 81: 7885-7889.

2. Donehower LA, Varmus HE. (1984) A mutant murine leukemia virus with a single missense codon in pol is defective in a function affecting integration. Proc. Natl. Acad. Sci. USA 81: 6461-6465.

3. Schwartzberg P, Colecilli J, Goff SP. (1984) Construction and analysis of deletion mutations in the pol gene of moloney murine leukemia virus: A new viral function required for productive infection. Cell 37: 1043-1052.

4. Carteau S, Hoffmann C, Bushman FD. (1998) Chromosome structure and HIV-1 cDNA integration: Centromeric alphoid repeats are a disfavored target. J. Virol. 72: 4005-4014.

5. Holman AG, Coffin JM. (2005) Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. Proc Natl Acad Sci U S A 102: 6103-6107.

6. Berry C, Hannenhalli S, Leipzig J, Bushman FD. (2006) Selection of target sites for mobile DNA integration in the human genome. PLoS Comput Biol 2(11): e157.

7. Farnet C, Bushman FD. (1997) HIV-1 cDNA integration: Requirement of HMG I(Y) protein for function of preintegration complexes in vitro. Cell 88: 1-20.

8. Chen H, Engelman A. (1998) The barrier-to-autointegration protein is a host factor for HIV type 1 integration. Proc. Natl. Acad. Sci. USA 95: 15270-15274.

9. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP. (1994) Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science 266(5193): 2002-6.

10. Christ F, Thys W, De Rijck J, Gijsbers R, Albanese A, et al. (2008) Transportin-SR2 imports HIV into the nucleus. Curr Biol 18(16): 1192-1202.

11. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, et al. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J. Biol. Chem. 278: 372-381.

12. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110(4): 521-529.

13. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, et al. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol 2(8): E234.

14. Wu X, Li Y, Crise B, Burgess SM. (2003) Transcription start regions in the human genome are favored targets for MLV integration. Science 300(5626): 1749-1751.

15. Poeschla EM. (2008) Integrase, LEDGF/p75 and HIV replication. Cell Mol Life Sci 65(9): 1403-1424.

16. Engelman A, Cherepanov P. (2008) The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. PLoS Pathog 4(3): e1000046.

17. Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, et al. (2005) Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. J Biol Chem 280(27): 25517-23.

18. Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, et al. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J. Biol. Chem. 278: 33528-33539.

19. Llano M, Vanegas M, Fregoso O, Saenz D, Chung S, et al. (2004) LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. J. Virol. 78: 9524-9537.

20. Llano M, Delgado S, Vanegas M, Poeschla EM. (2004) LEDGF/p75 prevents proteasomal degradation of HIV-1 integrase. J. Biol. Chem. 279: 55570-55577.

21. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, et al. (2006) An essential role for LEDGF/p75 in HIV integration. Science.

22. Marshall H, Ronen K, Berry C, Llano M, Sutherland H, et al. (2007) Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. PLoS One 2(12): e1340.

23. Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, et al. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev 21(14): 1767-1778.

24. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, et al. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med 11(12): 1287-1289.

25. Silvers R, Smith JA, Schowalter M, Litwin S, Liang Z, et al. (2009) Modification of integration site preferences of an HIV-1-based vector by expression of a novel synthetic protein. Hum Gene Ther .

26. Gijsbers R, Ronen K, Vets S, Malani N, De Rijck J, et al. (2010) LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. Mol Ther .

27. Konig R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, et al. (2008) Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell 135(1): 49-60.

28. Xie J, Beickman K, Otte E, Rymond BC. (1998) Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. EMBO J 17(10): 2938-2946.

29. Ookata K, Hisanaga S, Sugita M, Okuyama A, Murofushi H, et al. (1997) MAP4 is the in vivo substrate for CDC2 kinase in HeLa cells: Identification of an M-phase specific and a cell cycle-independent phosphorylation site in MAP4. Biochemistry 36(50): 15873-15883.

30. Chang W, Gruber D, Chari S, Kitazawa H, Hamazumi Y, et al. (2001) Phosphorylation of MAP4 affects microtubule properties and cell cycle progression. J Cell Sci 114(Pt 15): 2879-2887.

31. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, et al. (2008) Identification of host proteins required for HIV infection through a functional

genomic screen. Science 319(5865): 921-926. 10.1126/science.1152725.

32. Sun XJ, Wei J, Wu XY, Hu M, Wang L, et al. (2005) Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. J Biol Chem 280(42): 35261-35271.

33. Edmunds JW, Mahadevan LC, Clayton AL. (2008) Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J 27(2): 406-420.

34. Yoh SM, Lucas JS, Jones KA. (2008) The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev 22(24): 3422-3434.

35. Follenzi A, Ailes LE, Bakovic S, Gueuna M, Naldini L. (2000) Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. Nat. Genetics 25: 217-222.

36. Naldini L, Blomer U, Gallay P, Ory D, Mulligan R, et al. (1996) In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science 272: 263-267.

37. O'Doherty U, Swiggard WJ, Jeyakumar D, McGain D, Malim MH. (2002) A sensitive, quantitative assay for human immunodeficiency virus type 1 integration. J. Virol. 76: 10942-10950.

38. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. (2007) HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res 17(8): 1186-1194.

39. Jackson AL, Linsley PS. (2010) Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. Nat Rev Drug Discov 9(1): 57-67.

40. Xie P, Tian C, An L, Nie J, Lu K, et al. (2008) Histone methyltransferase protein SETD2 interacts with p53 and selectively regulates its downstream genes. Cell Signal 20(9): 1671-1678.

41. Tang Z, Li B, Bharadwaj R, Zhu H, Ozkan E, et al. (2001) APC2 cullin protein and APC11 RING protein comprise the minimal ubiquitin ligase module of the anaphase-promoting complex. Mol Biol Cell 12(12): 3839-3851.

42. Kataoka N, Bachorik JL, Dreyfuss G. (1999) Transportin-SR, a nuclear import receptor for SR proteins. J Cell Biol 145(6): 1145-1152.

43. Krishnan L, Matreyek KA, Oztop I, Lee K, Tipper CH, et al. (2010) The requirement for cellular transportin 3 (TNPO3 or TRN-SR2) during infection maps to human immunodeficiency virus type 1 capsid and not integrase. J Virol 84(1): 397-406.

44. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950): 289-293.

45. Ciuffi A, Mitchell RS, Hoffmann C, Leipzig J, Shinn P, et al. (2006) Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts. Mol Ther 13(2): 366-373.

46. Brady T, Agosto LM, Malani N, Berry CC, O'Doherty U, et al. (2009) HIV integration site distributions in resting and activated CD4+ T cells infected in culture. AIDS 23(12): 1461-1471.

CHAPTER 6 – THE EFFECT OF RALTEGRAVIR TREATMENT ON HIV INTEGRATION SITE SELECTION

6.1 Abstract

HIV shows a preference for integration in transcription units and gene-dense parts of the genome. This preference is thought to be mediated by interactions of viral components of the pre-integration complex with host chromatin-bound factors that act as integration tethers. This has been shown with one host factor, LEDGF/p75, which interacts with integrase proteins of HIV and other lentiviruses, and whose depletion impairs integration and alters its genomic distribution. It is thought that additional factors – either other tethering proteins or passive chromatin accessibility – also contribute to HIV integration targeting. A recent approach to identifying novel host factors has been to study factors that, like LEDGF/p75, are necessary for efficient infection, and determining the effect of their depletion on HIV integration targeting. We tested the effect on integration targeting of impairing infection by pharmacological means. Cells were treated with reverse transcriptase inhibitors AZT or Nevirapine, or the integrase inhibitor Raltegravir. We found that Raltegravir treatment altered the genomic distribution of integration sites, causing a shift to less gene-dense and GC-rich parts of the genome. Other inhibitors had no significant effect. These changes resemble those observed in HIV cofactor knockdown, leading us to speculate that a similar mechanism may be involved. We

tested the idea that retargeting was related to a delay in the timing of integration, but found no conclusive evidence in support of this.

6.2 Introduction

Integration is a necessary step in the lifecycle of retroviruses such as HIV, enabling the virus to establish life-long infection and form a latent reservoir. The integration reaction is mediated by the viral enzyme integrase (IN) [1-3]. IN mediates two reactions: terminal cleavage, whereby the last two nucleotides are removed from the end of each LTR of the reverse-transcribed genome; and strand transfer, where the target DNA is nicked once on each stand and the two recessed 3' hydroxyl groups exposed by terminal cleavage are inserted in a concerted fashion [4-7]. HIV IN is a 288 amino-acid protein in the RNaseH superfamily, composed of 3 structural domains. The central catalytic core domain (aa 51-212) contains an RNaseH fold found in many DNA and RNA modifying enzymes [8, 9]. Three acidic residues, D-DX_{35}-E, referred to as the catalytic triad, coordinate two divalent metal ions necessary for catalysis [10, 11].

As an essential viral enzyme without a cellular counterpart, IN makes a good target for the development of antiretrovirals. In 2007, the first integrase inhibitor, Raltegravir, developed by Merck, was approved by the FDA, and has since had great success in the clinic [12]. Raltegravir belongs to a class of compounds called diketo acids, thought to function by chelating the metal cations at the IN active site [13, 14]. This class of inhibitors specifically inhibits strand transfer, with much weaker inhibition of terminal cleavage [13]. It is thought that Raltegravir binds selectively to

IN in complex with viral DNA, and prevents the complex from binding to target DNA [15]. Cytoplasmic pre-integration complexes (PICs) from diketo-acid treated cells showed reduced in vitro strand transfer activity, suggesting that Raltegravir may bind the IN-viral DNA complex before entry into the nucleus [13, 16].

Integration events are non-randomly distributed in the genome of the infected cell, with retroviruses showing genus-specific preferences for genomic features [17]. HIV and related lentivruses show a preference for integration in active transcription units and relatively gene-rich regions [18]. This non-random genomic distribution is thought to effect optimal viral gene expression – it has been shown that transcriptionally silent HIV proviruses tend to show more integration in normally disfavored regions such as gene deserts and centromeric alphoid repeats [19]. The integration preferences of lentivruses such as HIV have been shown to be mediated by a chromatin-associated host protein, LEDGF/p75, which interacts with IN, tethering integration to chromatin, probably within the bodies of transcription units [20-22]. However, the genomic distribution of lentiviral integration sites in the absence of LEDGF/p75 is not fully random, suggesting other influences may also be important. Suggestions of such influences have included effects of the cell cycle [23], additional tethering factors (discussed in Chapter 5 of this dissertation) and passive chromatin accessibility [24].

In this chapter, we present data that pharmacological inhibition of HIV integration by Raltegravir also alters the genomic distribution of integration sites. 293T or Jurkat cells were treated with concentrations of Raltegravir that permitted

low levels of infection. Analysis of the genomic distribution of resultant integration sites showed that while IN's weak sequence specificity was not affected, integration sites from treated cells were found in less gene-dense and GC-rich regions. This was not the case when infection was inhibited with reverse transcriptase inhibitors AZT or Nevirapine, nor when untreated cells were infected at a lower MOI. While the mechanism of this effect remains unclear, it does not appear to result from a delay in integration kinetics in the presence of Raltegravir.

These findings are provocative in our developing understanding of the factors shaping retroviral integration targeting, and also of the effect of this new class of antiretrovirals on aspects of HIV replication in treated patients.

6.3 Materials and Methods

Cell culture

293T cells were cultured in Dulbecco's Modified Eagle Medium with Glutamax (Invitrogen), 10% FBS (Sigma) and 50μg/ml gentamycin (Sigma). Jurkat cells were cultured in RPMI (Invitrogen), 10% FBS (Sigma) and 50μg/ml gentamycin (Sigma).

Virion production and infections

VSV-G pseudotyped HIV vector particles were produced by Lipofectamine transfection of 293T cells with p156RRLsin-PPTCMVGFPWPRE [25], the packaging construct pCMVdeltaR9 [26], and the vesicular stomatitis virus G-producing pMD.G construct. Viral supernatant was harvested 38 hours after transfection, filtered through 0.22μm filters, concentrated by filtration through a

Centricon, treated with DNase I, and stored frozen at -80°C. HIV titer was quantified by p24 ELISA.

293T cells were seeded in 12-well plates, 300,000 cells per well, 6-8h before infection. Jurkat cells were aliquoted into 24-well plates, 500,000 cells per well, at the time of infection. For 'high MOI' infections, 293T cells were infected with 60ng p24 per well, Jurkats with 100ng p24 per well. For 'low MOI' infections, 300,000 293T cells were infected with 20ng p24 per well. The infection medium contained the appropriate culture medium, virus and DMSO, AZT, Nevirapine or Raltegravir at the concentrations stated. Each condition was conducted in quadruplicate. Infections were allowed to proceed overnight, then medium was replaced with fresh medium containing inhibitor or DMSO as appropriate. Cells treated with AZT and Nevirapine were harvested 48h after infection. Cells treated with Raltegravir were passaged for 2 weeks after infection to dilute unintegrated viral genomes. Cells treated with DMSO and infected at lower MOI were harvested after 48h, and cells treated with DMSO and infected at higher MOI were passaged for 2 weeks before harvesting. When passaging, inhibitor concentrations used at the time of infection were maintained.

For the time-course experiment, 250,000 293T cells were seeded per well of 12-well plate, seeding 4 wells per time-point. The following morning, cells were infected with 60ng p24 and 2.5ul DEAE-dextran per well. Each condition was conducted in quadruplicate. Infection was allowed to proceed for 2hr. The 0h time-point was harvested immediately after removal of the infection mix. For wells containing longer time-points, infection mix was removed and replaced with DMEM

with 10% FBS and 50μg/ml gentamycin containing DMSO or 10nM Raltegravir as appropriate. The same medium was left on until cells were harvested, at the time-point stated.

Integration site analysis

Integration sites were isolated and sequenced by ligation-mediated PCR essentially as described previously [27]. Genomic DNA was extracted from infected cells using the Qiagen DNeasy tissue extraction kit. Up to 2μg of DNA from each infection was digested overnight using MseI. This was followed by digestion to prevent amplification of internal viral fragments (from the 5' LTR) and plasmid backbone with SacI and DpnI. Linkers were then ligated onto digested products (oligonucleotide sequences listed below) and nested PCR performed from ligation products. Nested PCR primers contained 8 nt barcode sequences between the sequencing primer and LTR-binding portions. These enabled pooling of all PCR products into one sequencing reaction and subsequent separation of sequences by decoding the barcodes. Samples were sequenced on the Roche 454 GS-FLX instrument at the University of Pennsylvania.

Integration sites were judged to be authentic if the sequences had a best unique hit when aligned to the murine (mm8 draft) using BLAT, and the alignment began within 3bp of the viral LTR end and had >98% sequence identity. Detailed statistical methods are described in [28].

To control for possible biases in isolating integration sites due to restriction enzyme sequence distribution, three or ten matched random controls were

computationally generated for each experimental integration site that were the same distance from the closest MseI restriction site as the experimental site.

Integration site counts in various genomic annotations were compared with matched random controls by the Fisher's exact test. Additionally, multiple regression models for integration intensity were applied, as described in [28]. Analysis was carried out in the R statistical package (http://www.r-project.org).

Q-PCR for integrated proviruses

A two-step Alu-PCR assay was used, described in reference [29]. This consists of an intitial round of amplification from genomic Alu repeats to viral *Gag* sequence, followed by Q-PCR specific for an amplicon in R-U5 with molecular beacon probes on an Applied Biosystems 7500 Fast Realtime PCR instrument. Each sample was run in triplicate. In order to standardize values across QPCR plates, samples were run on multiple plates as external controls. Values for each plate were set to these control values.

6.4 Results

Isolation of HIV integration sites

293T or Jurkat cells were infected with a VSV-G pseudotyped GFP-encoding HIV vector in the presence of Raltegravir or DMSO as a control. As additional controls, 293T cells were infected in the presence of AZT or Nevirapine, and infection with DMSO was repeated with a lower virus inoculum. AZT- and Nevirapine-treated cells were harvested 48 hours after infection. Raltegravir-treated cells were found to accumulate 2-LTR circles, consistent with Raltegravir function

[13], which contaminated integration site amplification. These cells were therefore passaged for 14 days before harvesting. The lower MOI infection was harvested 48h post-infection, the higher after 2 weeks of passaging. The conditions used are shown in Table 6-1.

Integration sites were isolated and amplified from genomic DNA essentially as previously described [30]. The numbers of unique integration sites passing quality control are shown in Table 6-1. For reference, values for computationally generated random sites are shown. Random sites were generated that are matched to each dataset; Table 6-1 shows data from sites matched to the 293T DMSO high MOI dataset (see Materials and Methods for explanation).

Consensus sequence preference at HIV integration sites is preserved in the presence of antiretrovirals

Retroviral integrases show weak consensus sequences at the site of integration, consisting of an inverted repeat that varies between viruses [28, 31-35]. This consensus is a property of the integrase protein, with the symmetry of the inverted repeat thought to arise from IN binding as a dimer. In the case of HIV IN, the consensus is TDG↓GTWACCHA, where the arrow represents the site of integration [32, 35]. We therefore verified that in the presence of antiretrovirals, the target sequence preference of IN was not affected. Figure 6-1 shows that target site consensus sequences from the different treatments did not vary and were consistent with the published sequence. This indicates that the integration sites isolated were bona fide integration events, catalyzed by HIV IN.

Data set	% GFP positive (FACS)	Integration sites	% in genes	Average number genes in 1Mb	Average GC % in 1Mb
293T DMSO high MOI	89.6	871	72.4	20.7	43.7
293T DMSO low MOI	40.9	300	73.7	20.0	43.5
293T AZT 300nM	25.5	293	71.7	18.8	42.9
293T Nevirapine 100nM	11.5	415	71.6	22.0	44.3
293T Nevirapine 300nM	24.1	114	78.9	18.2	43.5
293T Raltegravir 10nM	31.5	316	74.7	17.4**	42.4**
293T Raltegravir 25nM	13.2	68	73.5	16.9*	42.2*
Jurkat DMSO	98.6	1890	73.8	19.6	43.5
Jukat Raltegravir 5nM	42.3	1225	75.4	15.8***	41.7***
Jurkat Raltegravir 10nM	36.3	1218	76.0	15.2***	41.4***
Jurkat Raltegravir 100nM	20.3	683	73.9	14.3***	41.0***
Random		5670	35.7***	9.95***	40.3

Table 6-1. Integration sets generated in this study and their genomic distributions. Cells were infected under the conditions shown. 'Random' is a set of computationally generated random sites in the genome, matched to the 'DMSO high MOI' set in terms of the distance of each site to the nearest MseI restriction site (see Methods for details). Asterisks denote statistical significance by the Fisher's exact test or Mann-Whitney U test as appropriate in the comparison with 293T DMSO high (for 293T sets) or Jurkat DMSO (for Jurkat sets). * $p<0.05$, ** $p<0.01$, ***$p<0.0001$.

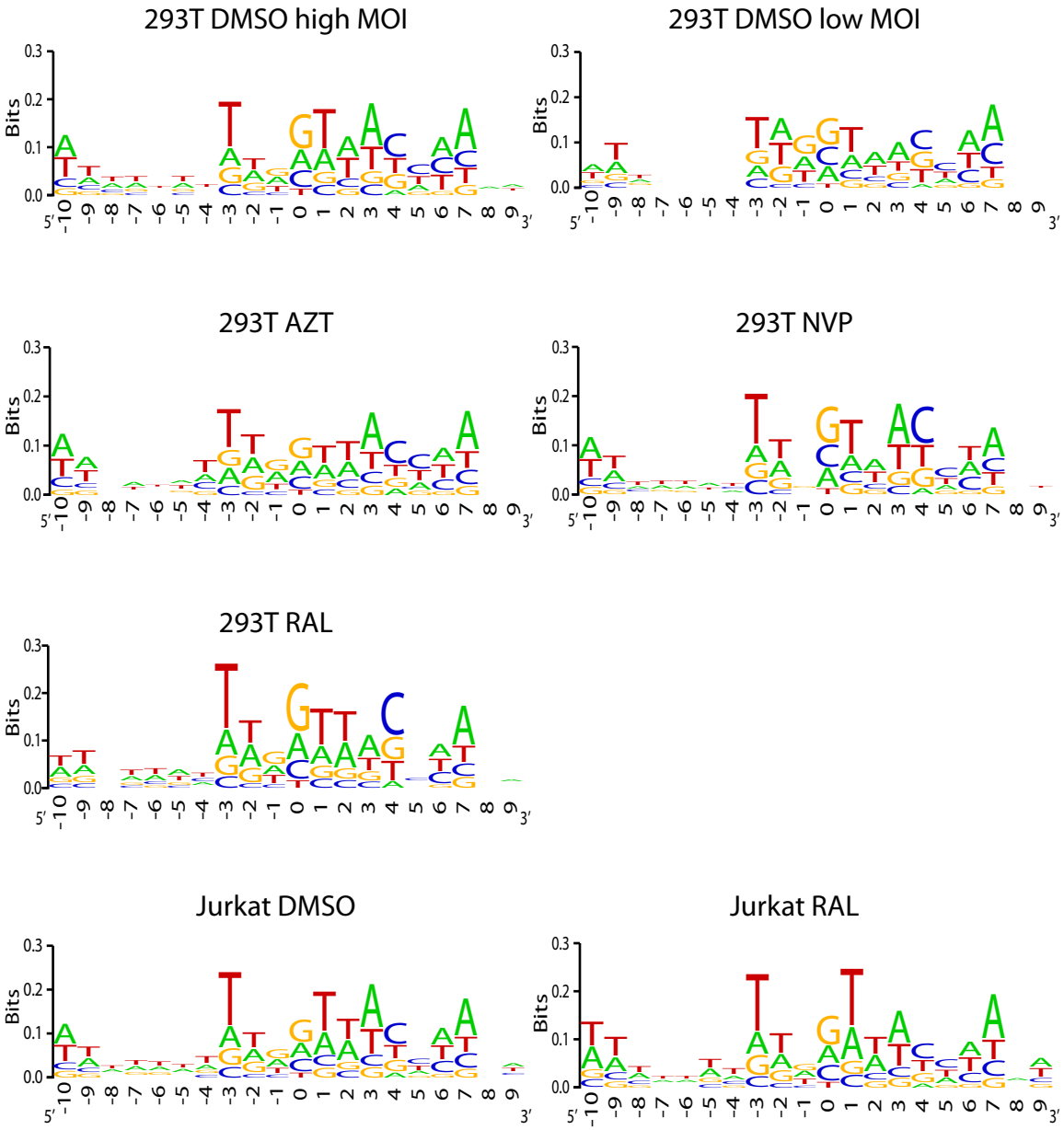

Figure 6-1. HIV integrase consensus sequence is not affected by antiretroviral treatment. The 20bp sequences immediately surrounding integration sites from each treatment were aligned to generate a consensus using the Weblogo program (weblogo.berkeley.edu). The x-axis represents the base position relative to the site of integration (between position -1 and 0). The y-axis represents the bits of information contained at each position (the maximal value, if a position only ever had one of the 4 bases, would be 2 bits).

HIV integration site selection is altered by Raltegravir treatment

HIV integration normally favors the bodies of active transcription units, and regions of the genome with a high gene density [18]. We therefore started by examining the correlation of integration sites with these features. Table 6-1 summarizes the proportions of integration sites in each treatment falling within transcription units annotated by the RefSeq gene call. It can be seen that, consistent with published data, HIV integration is roughly two-fold enriched in transcription units over random. No statistically significant differences were observed between cell types, MOIs or antiretroviral treatments.

We next examined the correlation between integration frequency and gene density. The average number of genes within a 1Mb window around sites from each treatment is shown in Table 6-1. It can be seen that sites from cells infected at different MOIs did not differ in the gene density of the surrounding DNA. Reverse transcriptase inhibitors AZT and Nevirapine also had no effect on the gene density surrounding HIV integration sites. However, Raltegravir treatment caused a small but statistically significant drop in the average gene density in a 1Mb window around HIV integration sites, an effect which was reproduced in both 293T and Jurkat cell lines: in 293T cells, sites from control cells had on average 20.7 genes in the surrounding 1Mb, which was reduced to 17.4 an 16.9 in 10nM and 25nM Raltegravir; sites from Jurkat cells similarly dropped from 19.6 genes per Mb to 15.8 and 15.2 at 5nM and 10nM respectively. Since, based on Table 6-1, it appeared that different dosages of each antiretroviral showed the same trend, sites were pooled across

concentrations in subsequent analysis. Figure 6-2 shows pairwise comparisons of the correlation between integration frequency and gene density in control and drug-treated cells. These graphs were made by combining the two datasets to be compared, binning sites into 10 bins of increasing gene density with equal numbers of sites in each bin, and plotting the proportion of each dataset falling within that bin (described in detail in reference [28]). This enables a sensitive analysis of differences between sets. This analysis confirmed that the distribution of integration sites from cells infected at different MOIs or in the presence of RT inhibitors AZT or Nevirapine did not differ significantly from sites from control infections. However, integration sites from Raltegravir-treated cells were more commonly found in bins of lower gene density than sites from control cells. The effect was more pronounced in Jurkat cells, though these sets also contained more integration sites than those from 293T cells.

A number of genomic features correlate in the genome, for example gene-rich regions also tend to have short genes, short introns and be high in GC content. We therefore also examined the relationship between integration frequency and GC content within a 1Mb window around each site. The effects on GC content are shown in Table 6-1, and were similar to those for gene density. While DMSO treatment, MOI and RT inhibition had no effect on the average GC content surrounding integration sites, regions around sites from Raltegravir-treated cells showed a small but statistically significant and reproducible drop in GC content (in 293Ts from 43.7% to 42.4% or 42.2%, in Jurkats from 43.5% to 417% or 41.4%). Sensitive

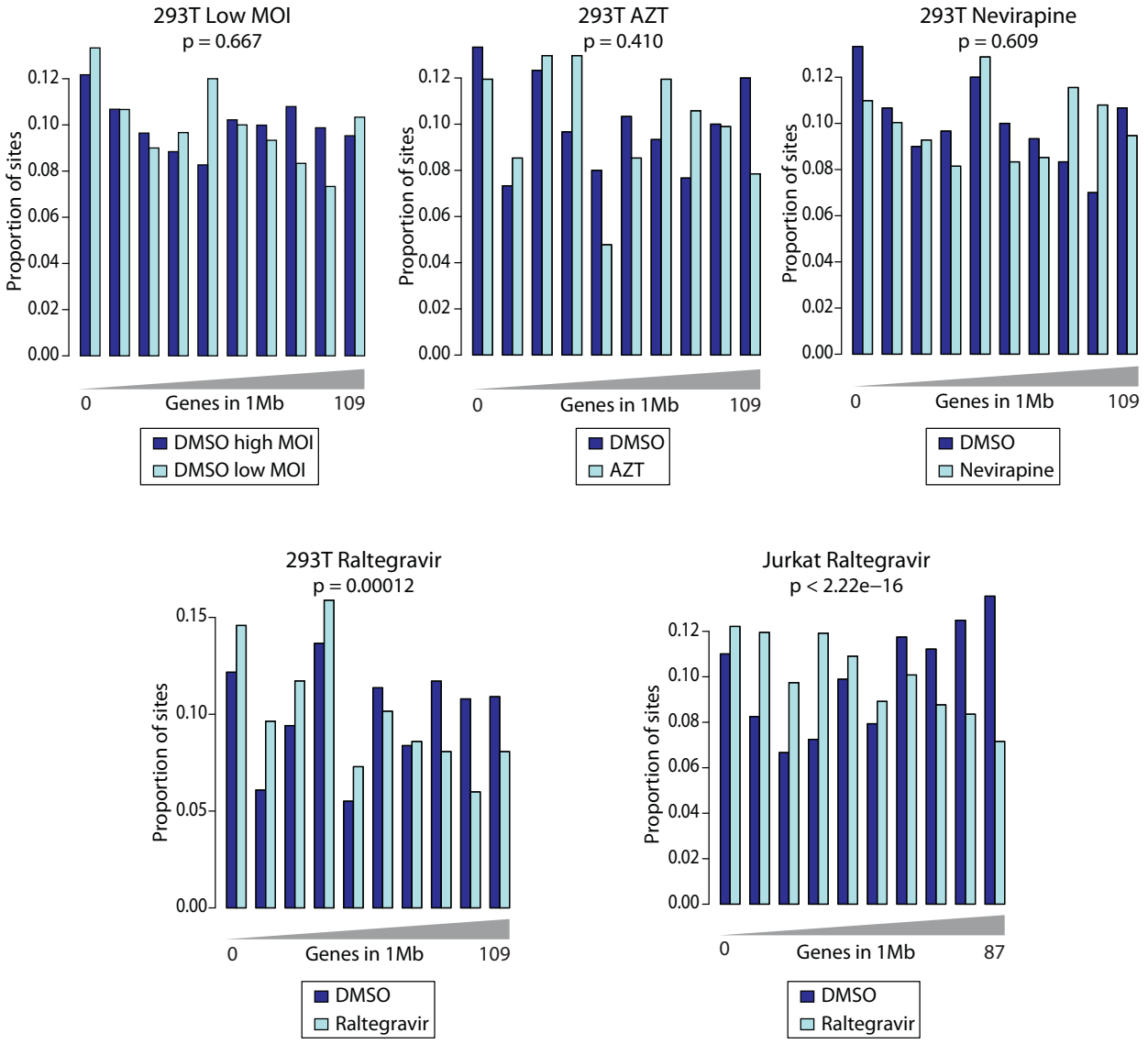

Figure 6-2. The effect of antiretroviral treatment on integration with respect to gene density. HIV vector integration sites from cells subjected to each of the treatments shown were isolated and mapped onto the genome. The number of Refseq genes in a 1Mb window around each site was calculated. Sites from control and treated sets were combined and split into ten bins of increasing gene density with equal numbers of sites in each bin. The proportion of each set found in each bin is shown. P values were determined using the likelihood ratio statistic for the logistic regression model. 'DMSO' in 293T panels refers to high MOI infection of DMSO-treated cells.

graphical pairwise comparisons between sets are shown in Figure 6-3. A shift can be seen in the distribution of sites to regions of lower GC content following Raltegravir treatment in both cell types.

Raltegravir does not influence the kinetics of HIV integration in the cell

We were surprised by the effects of Raltegravir treatment on integration site selection, since prevalent models of integration targeting have incorporated host tethering factors and variation in chromatin accessibility [24, 36], which we would not expect to be affected by Raltegravir. We wondered if the observed effect could be explained by a kinetic effect on integration – that partial inhibition of Raltegravir as was performed in this experiment might act to alter the timing of successful integration events, and thus alter the time available for the unintegrated PIC to traffic through the nucleus. It is known that chromatin folds into higher order structures in the nucleus, with particular genomic features clustering in particular places [37]. It could be imagined that the PIC might need to travel through the nucleus to its final site of integration, and that the time available to complete this journey could affect the destination reached. We expected that Raltegravir treatment might slow the integration reaction, lengthening the time of PIC nuclear trafficking.

To establish whether integration was slowed by Raltegravir treatment, we conducted a time-course of infection with and without Raltegravir and measured by quantitative PCR integrated proviral copies at intervals following infection. 293T cells were infected for 2 hours with VSV-G pseudotyped virus in the presence of 10nM Raltegravir or DMSO as control. Virus was then removed and replaced with

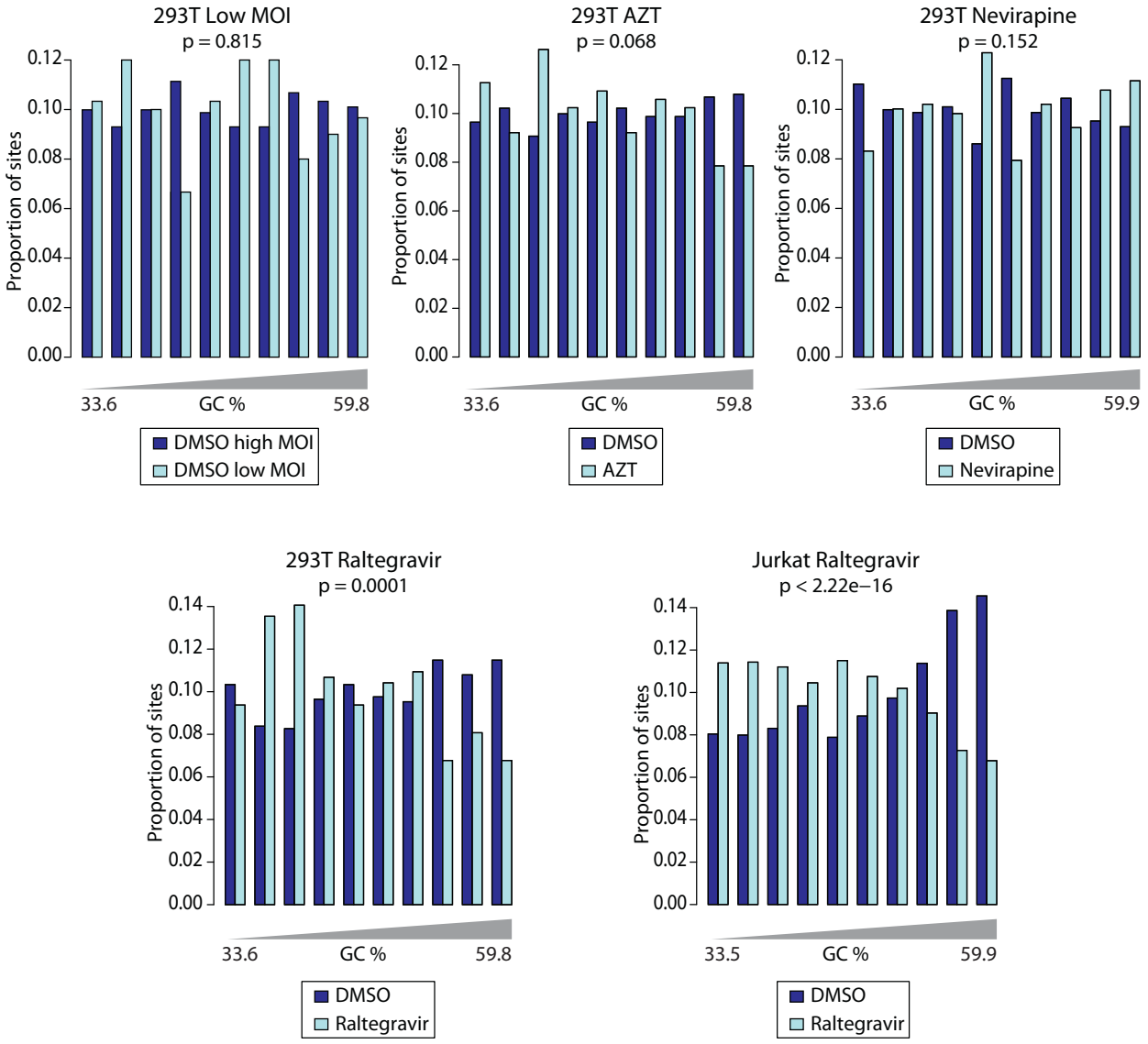

Figure 6-3. The effect of antiretroviral treatment on integration with respect to GC content. HIV vector integration sites from cells subjected to each of the treatments shown were isolated and mapped onto the genome. The GC content in a 1Mb window around each site was calculated. Sites from control and treated sets were combined and split into ten bins of increasing gene density with equal numbers of sites in each bin. The proportion of each set found in each bin is shown. P values were determined using the likelihood ratio statistic for the logistic regression model. 'DMSO' in 293T panels refers to high MOI infection of DMSO-treated cells.

medium containing 10nM Raltegravir or DMSO as appropriate for the remainder of the time-course. Cells were harvested at intervals over the next 72 hours and an Alu-repeat-based quantitative PCR assay performed to quantify integrated proviruses. The results are shown in Figure 6-4. It can be seen that starting around 6h post-infection, the number of proviruses started to increase in both control and treated cells. In both conditions, the level of integrated proviruses increased until a plateau was reached. Treatment with Raltegravir reduced this plateau, as expected. As an indication of the timing of integration in each condition, we calculated the time taken to synthesize half of the maximal quantity of proviruses for each condition ($t_{1/2max}$). Based on this analysis, integration had reached half its maximal value at 16.9h in control cells and 13.9h in Raltegravir-treated cells. Further repetitions of the experiment and more time-points between 12h and 24h would be required to establish whether there was any statistically significant difference between the two treatments, though these results do not suggest any large effects, and certainly provide no evidence for retardation of integration.

6.5 Discussion

HIV integration events are distributed in the host genome with characteristic preferences for transcription units and gene-dense regions. It is known that LEDGF/p75 contributes to this distribution, but it is possible that other factors play a role. Here we present data that pharmacological inhibition of HIV integrase by Raltegravir leads to a redistribution of integration sites into less gene-dense and GC-rich regions. The changes were modest but significant and reproducible in both

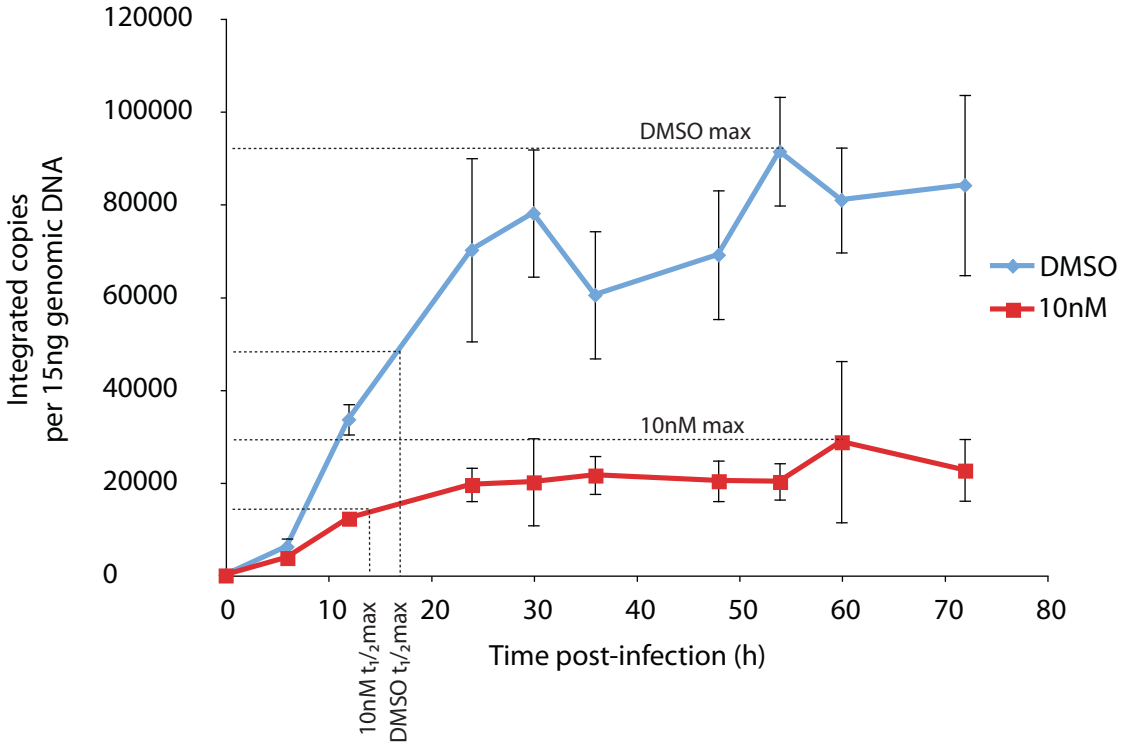

Figure 6-4. The effect of Raltegravir treatment on HIV integration kinetics. 293T cells were infected with HIV in the presence of DMSO or 10nM Raltegravir as shown. Cells were harvested at intervals after infection and the number of integrated proviruses quantified by Q-PCR.

293Ts and Jurkats, and absent in sites from infection at a lower MOI or inhibited at reverse transcription. In trying to explain this observation, we investigated one possible explanation – that partial inhibition with Raltegravir altered the timing of integration and that alterations in PIC nuclear trafficking time might affect the final location of integration. A preliminary experiment did not provide strong evidence in favor of this hypothesis, though it suggested that in treated cells successful integration events, if different, take place earlier. Further experiments with time-points at shorter intervals would be required to reach a definitive conclusion. If timing were reproducibly altered by Raltegravir treatment, it could be interesting to analyze the distribution of integration sites at various times after infection to see if certain sites were earlier targets.

Another possibility is that Raltegravir induces cellular cytotoxicity, which leads to either changes in expression of host integration site determinants or global changes in chromatin structure that influence its accessibility to the PIC. This possibility could be ruled out by analyzing integration sites from cells infected with Raltegravir-resistant virus in the presence and absence of drug, analogous to siRNA-resistant expression commonly employed in siRNA experiments, as described in Chapter 5.

Finally, it has recently been found that type I interferons can be induced by cytoplasmic DNA through a TLR-independent mechanism named the Interferon-stimulatory DNA (ISD) response [38, 39]. It has been suggested that cDNA from reverse transcription of retroviruses and endogenous retroelements may activate the

ISD response [40-42]. Perhaps inhibition of viral integration by Raltegravir and the resultant accumulation of viral cDNA and 2-LTR circles induces an innate immune response such as the ISD that leads to changes in gene expression or chromatin structure that affect integration site distribution. However, whether the ISD response is inducible by retroviral infection in the 293T and Jurkat cell lines employed here is unclear [39]. A potential experiment to test this idea would be to monitor the expression of interferon-inducible genes in cells treated with Raltegravir. Additionally, cells could be treated with interferon-β and integration sites analyzed for recapitulation of the effects observed here. The similarity between the effects of Raltegravir and the effects of siRNA treatment described in Chapter 5 are striking, and suggestive of some common underlying mechanism. Perhaps induction of an interferon response, either in response to the accumulation of viral replication intermediates through the ISD, or in response to siRNA treatment, causes chromatin or gene expression changes that result in the effects documented here and in the preceding chapter.

The mechanism of the observations described here therefore remains unclear. However, it is of interest that an antiretroviral drug in clinical use appears to alter the genomic distribution of HIV integration sites. An important question would be whether these changes have any effect on HIV replication. Lewinski *et al.* [19] reported that the positioning of proviruses bears on their transcriptional activity: transcriptionally silent proviruses were more often found in gene deserts and centromeric repeats. Therefore, one question would be whether the alteration of

integration sites by Raltegravir treatment contributes to its antiviral role by impairing proviral expression.

6.6 References

1. Donehower LA, Varmus HE. (1984) A mutant murine leukemia virus with a single missense codon in pol is defective in a function affecting integration. Proc. Natl. Acad. Sci. USA 81: 6461-6465.

2. Schwartzberg P, Colecilli J, Goff SP. (1984) Construction and analysis of deletion mutations in the pol gene of moloney murine leukemia virus: A new viral function required for productive infection. Cell 37: 1043-1052.

3. Panganiban AT, Temin HM. (1984) The retrovirus pol gene encodes a product required for DNA integration: Identification of a retrovirus int locus. Proc. Natl. Acad. Sci. USA 81: 7885-7889.

4. Brown PO, Bowerman B, Varmus HE, Bishop JM. (1989) Retroviral integration: Structure of the initial covalent complex and its precursor, and a role for the viral IN protein. Proc. Natl. Acad. Sci. U S A 86: 2525-2529.

5. Fujiwara T, Mizuuchi K. (1988) Retroviral DNA integration: Structure of an integration intermediate. Cell 54: 497-504.

6. Bushman FD, Fujiwara T, Craigie R. (1990) Retroviral DNA integration directed by HIV integration protein *in vitro*. Science 249: 1555-1558.

7. Craigie R, Fujiwara T, Bushman F. (1990) The IN protein of moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell 62: 829-837.

8. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, et al. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: Similarity to other polynucleotidyl transferases. Science 266: 1981-1986.

9. Nowotny M. (2009) Retroviral integrase superfamily: The structural perspective. EMBO Rep 10(2): 144-151.

10. Diamond TL, Bushman FD. (2006) Role of metal ions in catalysis by HIV integrase analyzed using a quantitative PCR disintegration assay. Nucleic Acids Res 34(21): 6116-25.

11. Gao K, Wong S, Bushman F. (2004) Metal binding by the D,DX35E motif of human immunodeficiency virus type 1 integrase: Selective rescue of cys substitutions by Mn2+ *in vitro*. J. Virol. 78(13): 6715-22.

12. Summa V, Petrocchi A, Bonelli F, Crescenzi B, Donghi M, et al. (2008) Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. J Med Chem . 10.1021/jm800245z.

13. Hazuda DJ, Felock P, Witmer M, Wolfe A, Stillmock K, et al. (2000) Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. Science 287: 646-650.

14. Grobler JA, Stillmock K, Hu B, Witmer M, Felock P, et al. (2002) Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes. Proc Natl Acad Sci U S A 99: 6661-6666.

15. Espeseth AS, Felock P, Wolfe A, Witmer M, Grobler J, et al. (2000) HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. Proc. Natl. Acad. Sci. U S A 97(21): 11244-9.

16. Pandey KK, Grandgenett DP. (2008) HIV-1 integrase strand transfer inhibitors: Novel insights into their mechanism of action. Retrovirology 2: 11-16.

17. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, et al. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol 2(8): E234.

18. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110(4): 521-529.

19. Lewinski M, Bisgrove D, Shinn P, Chen H, Verdin E, et al. (2005) Genome-wide analysis of chromosomal features repressing HIV transcription. J. Virol. .

20. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, et al. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med 11(12): 1287-1289.

21. Marshall H, Ronen K, Berry C, Llano M, Sutherland H, et al. (2007) Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. PLoS One 2(12): e1340.

22. Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, et al. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev 21(14): 1767-1778.

23. Ciuffi A, Mitchell RS, Hoffmann C, Leipzig J, Shinn P, et al. (2006) Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts. Mol Ther 13(2): 366-373.

24. Ciuffi A, Bushman FD. (2006) Retroviral DNA integration: HIV and the role of LEDGF/p75. Trends Genet. Epub.

25. Follenzi A, Ailes LE, Bakovic S, Gueuna M, Naldini L. (2000) Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. Nat. Genetics 25: 217-222.

26. Naldini L, Blomer U, Gallay P, Ory D, Mulligan R, et al. (1996) In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science 272: 263-267.

27. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. (2007) HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res 17(8): 1186-1194.

28. Berry C, Hannenhalli S, Leipzig J, Bushman FD. (2006) Selection of target sites for mobile DNA integration in the human genome. PLoS Comput Biol 2(11): e157.

29. O'Doherty U, Swiggard WJ, Jeyakumar D, McGain D, Malim MH. (2002) A sensitive, quantitative assay for human immunodeficiency virus type 1 integration. J. Virol. 76: 10942-10950.

30. Ciuffi A, Ronen K, Brady T, Malani N, Wang G, et al. (2008) Methods for integration site distribution analyses in animal cell genomes. Methods .

31. Pryciak PM, Sil A, Varmus HE. (1992) Retroviral integration into minichromosomes *in vitro*. EMBO J. 11: 291-303.

32. Stevens SW, Griffith JD. (1996) Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. J. Virol. 70: 6459-6462.

33. Carteau S, Hoffmann C, Bushman FD. (1998) Chromosome structure and HIV-1 cDNA integration: Centromeric alphoid repeats are a disfavored target. J. Virol.

72: 4005-4014.

34. Wu X, Li Y, Crise B, Burgess SM, Munroe DJ. (2005) Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. J. Virol. 79(5211-5214).

35. Holman AG, Coffin JM. (2005) Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. Proc Natl Acad Sci U S A 102: 6103-6107.

36. Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, et al. (2006) Retroviral DNA integration: Viral and cellular determinants of target-site selection. PLoS Pathog 2(6): e60.

37. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950): 289-293.

38. Ishii KJ, Coban C, Kato H, Takahashi K, Torii Y, et al. (2006) A toll-like receptor-independent antiviral response induced by double-stranded B-form DNA. Nat Immunol 7(1): 40-48.

39. Stetson DB, Medzhitov R. (2006) Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. Immunity 24(1): 93-103.

40. Stetson DB, Ko JS, Heidmann T, Medzhitov R. (2008) Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell 134(4): 587-598.

41. Beck-Engeser GB, Eilat D, Harrer T, Jack HM, Wabl M. (2009) Early onset of autoimmune disease by the retroviral integrase inhibitor raltegravir. Proc Natl Acad Sci U S A .

42. Yan N, Cherepanov P, Daigle JE, Engelman A, Lieberman J. (2009) The SET complex acts as a barrier to autointegration of HIV-1. PLoS Pathog 5(3): e1000327.

CHAPTER 7 – CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, I presented studies of the determinants and consequences of lentiviral integration in human and murine cells. Lentiviruses have evolved to preferentially integrate into transcription units and gene-dense regions in the genome. In Chapter 2 of this dissertation, I showed that much of this preference is determined by an interaction between the viral integrase protein and a host chromatin-associated protein, LEDGF/p75. I also reported that this preference could be modulated by variation in LEDGF/p75 expression within the biological range found in different cell types. However, integration was still not fully random in the absence of LEDGF/p75, with certain preferences unaffected and new preferences emerging, suggesting other factors in addition to LEDGF/p75 may be involved.

In Chapter 3 I presented data from cells expressing LEDGF/p75 fusion proteins with alternative chromatin-binding domains. I reported that a LEDGF/p75 fusion with the heterochromatin-binding protein CBX1 could retarget integration into normally disfavored sites: intergenic and pericentromeric regions, rich in marks of transcriptional repression. This argues that, while other factors may play a role in targeting integration to transcriptionally active genomic regions, LEDGF/p75 tethering appears to be central to targeting, and able to overcome potential barriers to integration in regions such as pericentric heterochromatin.

Candidate factors that may contribute to lentiviral targeting in addition to LEDGF/p75 were investigated in Chapters 5 and 6. These studies revealed that

knockdown of a number of cellular factors, or the induction of other changes to the cellular or nuclear environment, alter integration site selection. The mechanism remains unknown, but it appears that these effects differ from the effects of LEDGF/p75 depletion.

Clarifying the mechanism of the effects observed in Chapters 5 and 6 remains a goal of future experiments, for example by testing the idea that an innate immune response was elicited by the experimental conditions used. Analyzing integration sites in cells knocked down both for LEDGF/p75 and for the factors described in Chapter 5, or treating LEDGF/p75 knockdown cells with Raltegravir as in Chapter 6, would also be useful. Preliminary data from such an experiment suggest the effects described in Chapters 5 and 6 and the effects of LEDGF/p75 depletion described in Chapter 2 are indeed independent, and additive rather than epistatic (KR, data not shown). Additionally, mutants of HIV that escape restriction by the host factors TNPO3 and RANBP2 have been selected (Vineet KewalRamani, Greg Towers, personal communications). Integration site analysis with these viruses could be studied to test the proposed connection between the nuclear import pathway and integration targeting.

One important question arising from these studies concerns the implication of alterations in integration targeting for viral gene expression. It has been suggested that the lentiviral preference for integration in expressed genes maximizes proviral expression, and that integration into unfavorable genomic regions such as gene deserts impairs proviral expression, potentially contributing to latency [1, 2]. The

apparently efficient expression of vectors retargeted to heterochromatin in Chapter 3 is therefore surprising. In the future, epigenetic modifications of the integrated proviral DNA could be studied. The expression of reporter genes from vectors containing different promoters could also be measured, to address the possibility of promoter-specific effects. A complication in the experimental system used in Chapter 3 is the stable expression of the CBX1-LEDGF$_{325-530}$ fusion. It is possible that this protein interferes with the silencing mediated by endogenous CBX1 recruitment. The development of a system in which the fusion protein was only transiently expressed during integration, and subsequently switched off, would help test this idea – perhaps in the absence of the fusion protein proviral expression would be silenced as predicted. Another potentially interesting experiment would be to compare the integration site distributions of well and poorly expressed proviruses following integration in the presence of the CBX1 fusion, as was performed in reference [1]. This could enable us to ask more carefully if retargeted integration events lead to reduced proviral expression.

The study of integration retargeting presented in Chapter 3 also has implications for the use of lentiviral vectors in gene therapy. Previous instances of retroviral vector-mediated insertional activation and oncogenesis have led to concerns about the effect of integration in or near genes. The retargeting presented in Chapter 3 suggests that host integration cofactors could be manipulated to direct integration to parts of the genome considered to be 'safe'.

Whether natural patterns of lentiviral integration are a threat to vector safety is an issue addressed in Chapter 4. Here we analyzed the distribution of integration sites in mice treated for β-thalassemia with a lentiviral vector. We analyzed bone marrow after long-term reconstitution, asking if the distribution of integration sites showed evidence of growth stimulation by vector integration near genes involved in growth control. In contrast to a previous study in one human patient with the same vector (Cavazzana-Calvo *et al.* submitted), but in agreement with studies with other lentiviral vectors [3, 4], we found no evidence suggestive of insertional activation. However, we did observe that integration sites in cells having undergone long-term growth *in vivo*, or short-term culture *ex vivo*, were less frequently in transcription units. This is suggestive of negative selection against cells with proviruses in genes, as is seen with endogenous retroviruses [5-7]. Given the natural history of HIV infection, it is perhaps not surprising that integration of an HIV-based vector could place cells at a selective disadvantage – HIV-infected lymphocytes typically live a short time after infection [8] so there would be little pressure for HIV to evolve a benign integration pattern, as is seen in the yeast Ty retrotransposons for example [9-11].

Considering the findings of Chapters 3 and 4, one wonders where in the genome is a 'safe' integration site. Further analysis of the effect of lentiviral integration on host gene expression is warranted. Previous studies have shown that lentiviral integration can lead to both up- and down-regulation of nearby genes [12]. Do certain integration site distributions minimize the deregulation of host gene

expression? Transcriptional profiling of cells with different integration patterns could be conducted. LEDGF/p75 fusions present an attractive model in which the integration preference of a lentiviral vector can be modified and effects on host gene expression compared. Again, the current LEDGF/p75 fusion protein system generated in Chapter 3 is complicated by the other modifications of the cells, but refining the model could enable side-by side comparison.

The studies described in this dissertation contribute to our understanding of lentiviral integration, its determinants and some of its consequences. They help illuminate virus-host interactions, aiding in the identification of therapeutic targets, and contribute to the ongoing development and evaluation of the use of lentiviral vectors in gene therapy.

References

1. Lewinski M, Bisgrove D, Shinn P, Chen H, Verdin E, et al. (2005) Genome-wide analysis of chromosomal features repressing HIV transcription. J. Virol. .

2. Jordan A, Bisgrove D, Verdin E. (2003) HIV reporducibly establishes a latent infection after acute infection of T cells ni vitro. EMBO J. 22: 1868-1877.

3. Wang GP, Levine BL, Binder GK, Berry CC, Malani N, et al. (2009) Analysis of lentiviral vector integration in HIV+ study subjects receiving autologous infusions of gene modified CD4+ T cells. Mol Ther .

4. Cartier N, FAU - Hacein-Bey-Abina S, Hacein-Bey-Abina S, FAU - Bartholomae CC, Bartholomae CC, et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. - Science.2009 Nov 6;326(5954):818-23. (1095-9203 (Electronic)).

5. Smit AF. (1999) Interspersed repeats and other momentos of transposable elements in mammalian genomes. Curr Opin Genet Dev 9(6): 657-63.

6. Lander E. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921.

7. Brady T, Lee YN, Ronen K, Malani N, Berry CC, et al. (2009) Integration target site selection by a resurrected human endogenous retrovirus. Genes Dev 23(5): 633-642.

8. Wei X, Ghosh SK, Taylor ME, V J, Emini EA, et al. (1995) Viral dynamics in human immunodeficiency virus type 1 infection. Nature 373: 117-122.

9. Boeke JD, Devine SE. (1998) Yeast retrotransposons: Finding a nice quiet neighborhood. Cell 93(7): 1087-9.

10. Sandmeyer S. (2003) Integration by design. Proc. Natl. Acad. Sci. U S A 100: 5586-5588.

11. Xie W, Gai X, Zhu Y, Zappulla DC, Sternglanz R, et al. (2001) Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Mol Cell Biol 21(19): 6606-6614.

12. Maruggi G, Porcellini S, Facchini G, Perna SK, Cattoglio C, et al. (2009) Transcriptional enhancers induce insertional gene deregulation independently from the vector type and design. Mol Ther 17(5): 851-856.

Supplementary Table S4-1. Capture-recapture-based total population estimates

enzyme 1	enzyme 2 seen	unseen	**estimate m22=(m12.m21)/m11**
seen	m11	m12	
unseen	m21	**m22**	

MOUSE 31

mouse 31 uncultured	Nla				
Mse	yes	no	estimate m22	estimate total	
yes	10	20	30	16	**54**
no	8	m22			
	18				

mouse 31 methylcellulose	Nla		
Mse	yes	no	
yes	0	2	no overlap, can't estimate
no	3	m22	

mouse 31 all unique	Nla				
Mse	yes	no	estimate m22	estimate total	
yes	10	22	32	17.6	**57.6**
no	8	m22			
	18				

NOTE: 3/5 sites after methylcellulose are also found before culture

MOUSE 32

mouse 32 uncultured	Nla				
Mse	yes	no	estimate m22	estimate total	
yes	6	12	19	20	**48**
no	10	m22			
	16				

mouse 32 methylcellulose	Nla				
Mse	yes	no	estimate m22	estimate total	
yes	3	3	6	1	**8**
no	1	m22			
	4				

mouse 32 all unique	Nla				
Mse	yes	no	estimate m22	estimate total	
yes	6	12	19	20	**48**
no	10	m22			
	16				

NOTE: all 7 sites after methylcellulose are also found before culture

MOUSE 33.1

mouse 33.1 uncultured

Mse \ Nla	yes	no		estimate m22		estimate total	
yes	9	16	25		16		**50**
no	9	m22					
	18						

mouse 33.1 methylcellulose

Mse \ Nla	yes	no		estimate m22		estimate total	
yes	4	4	8		11		**30**
no	11	m22					
	15						

mouse 33.1 methylcellulose

Mse \ Nla	yes	no		estimate m22		estimate total	
yes	9	18	27		32		**75**
no	16	m22					
	25						

12/19 sites after MC are also found before culture

MOUSE 33.2

mouse 33.2 uncultured

Mse \ Nla	yes	no		estimate m22		estimate total	
yes	4	8	12		26		**51**
no	13	m22					
	17						

mouse 33.2 methylcellulose

Mse \ Nla	yes	no		estimate m22		estimate total	
yes	1	1	2		41		**84**
no	41	m22					
	42						

mouse 33.2 all unique

Mse \ Nla	yes	no		estimate m22		estimate total	
yes	4	7	11		82.25		**140.25**
no	47	m22					
	51						

NOTE: 9/43 sites after MC are also found before cultı

MOUSE 34

mouse 34 uncultured

Mse \ Nla	yes	no		estimate m22		estimate total	
yes	7	21	28		12		**44**
no	4	m22					
	11						

mouse 34 methylcellulose	Nla				
Mse	yes	no	estimate m22	estimate total	
yes	4	1	5	0.75	**8.75**
no	3	m22			
	7				

mouse 34 all unique	Nla				
Mse	yes	no	estimate m22	estimate total	
yes	7	21	28	15	**48**
no	5	m22			
	12				

NOTE: 7/8 sites after MC are also found before cultu

Appendix
CURRICULUM VITAE

EDUCATION

2010 (expected)	*University of Pennsylvania, Philadelphia, PA* **PhD, Cell and Molecular Biology** *Thesis title:* Lentiviral integration site selection: host determinants and consequences *Advisor:* Frederic Bushman
2005	*Churchill College, Cambridge University, UK* **BA, Natural Sciences (Pathology)** First Class

LABORATORY RESEARCH EXPERIENCE

2005-present	*University of Pennsylvania, Philadelphia, PA* ***Lentiviral integration site selection*** PhD candidate, Laboratory of Frederic Bushman
2004-2005	*Cambridge University, Cambridge, UK* ***The effect of HCMV infection on the cell cycle of myeloid cells*** Undergraduate student, Laboratory of John Sinclair
2004-2004	*University of Texas, Houston, TX* ***Inflammatory response to mycobacterial TDM in complement 5-deficient mice*** Summer student, Laboratory of Jeffrey Actor

ACADEMIC HONORS AND AWARDS

2009-2010	Funding for research in Africa (Health and Society in Africa seminar)
2008-2009	Virology Training Grant (NIH T32 AI-07324-17)
2008	Global Health Framework Fellowship
2006-2008	Cell & Molecular Biology Training Grant (NIH T32 07229-33)
2003-2005	Churchill College Scholarship

PUBLICATIONS

Ronen K, Negre O, Malani N, Denaro M, Gillet-Legrand B, Leboulch P, Down J, Payen E, Bushman FD. *Integration site distribution in mice following therapeutic gene transfer to treat β-thalassemia.* Submitted.

Gijsbers R, **Ronen K**, Vets S, Malani N, De Rijck J, McNeely M, Bushman FD, Debyser Z. *LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin.* Mol. Ther. advance online publication Jan 5, 2010

Brady T, Lee YN, **Ronen K**, Malani N, Berry CC, Bieniasz PD, Bushman FD. *Integration target site selection by a resurrected human endogenous retrovirus.* Genes Dev. 2009 Mar 1;23(5):633-42

Ciuffi A, **Ronen K**, Brady T, Malani N, Wang G, Berry CC, Bushman FD. *Methods for integration site distribution analyses in animal cell genomes.* Methods. 2009 Apr;47(4):261-8

Bushman FD, Hoffmann C, **Ronen K**, Malani N, Minkah N, Rose HM, Tebas P, Wang G. *Massively parallel pyrosequencing in HIV research.* AIDS. 2008 Jul 31;22(12):1411-5

Wang GP, Garrigue A, Ciuffi A, **Ronen K**, Leipzig J, Berry C, Lagresle-Peyrou C, Benjelloun F, Hacein-Bey-Abina S, Fischer A, Cavazzana-Calvo M, Bushman FD. *DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer.* Nucleic Acids Res. 2008 May;36(9):e49

Marshall HM*****, **Ronen K***, Berry C, Llano M, Sutherland H, Saenz D, Bickmore W, Poeschla E, Bushman FD. *Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting.* PLoS ONE. 2007 Dec 19;2(12):e1340 (*Equal contribution)

Borders CW, Courtney A, **Ronen K**, Pilar Laborde-Lahoz M, Guidry TV, Hwang SA, Olsen M, Hunter RL Jr, Hollmann TJ, Wetsel RA, Actor JK. *Requisite role for complement C5 and the C5a receptor in granulomatous response to mycobacterial glycolipid trehalose 6,6'-dimycolate.* Scand J Immunol. 2005 Aug;62(2):123-30.

CONFERENCE TALKS

2009	Cold Spring Harbor Retroviruses Meeting ***Novel host factors in HIV infection: effects on integration site selection***
2008	Mid-Atlantic Transposon Meeting ***Lentiviral integration site selection: LEDGF and histone modifications***

PROFESSIONAL ACTIVITIES

2009-present	Reviewer: Journal of Virology
2008-2009	Teaching: rotation student supervisor