A Hybrid Lentivirus-Transposon Vector for Safer Gene Therapy

Conrad Vink

Institute of Child Health

University College London

A thesis submitted for the degree of Doctor of Philosophy

June 2009

Declaration

I, Conrad Vink, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

1. Abstract

Gene therapy vectors based on the HIV-1 lentivirus are an attractive option for clinical applications because they enter a broad range of target cells efficiently and deliver stable gene expression through integration into host chromosomes. However, lentiviral vectors are known to integrate preferentially within actively transcribing genes. Leukaemia-like expansions observed in gene therapy trials using gammaretroviral vectors and are thought to have been caused by disruption of host proto-oncogenes at or close to the vector integration site, and the safety of these vectors may be related to the pattern of vector integration with respect to genes. The safety of integrating gammaretroviral and lentiviral vectors is therefore a significant concern with respect to their use in gene therapy. In this study, this problem was addressed by developing a novel hybrid vector which combines the efficient cell and nuclear entry properties of lentiviral vectors with chromosomal integration by the Sleeping Beauty transposase. Unlike the HIV-1 integrase enyme, Sleeping Beauty transposase does not exhibit a preference for integration within active genes. Non-integrating lentiviral vectors were developed to carry Sleeping Beauty transposon and transposase expression cassettes. These were able to deliver transient transposase expression to target cells, and episomal lentiviral DNA was found to be a suitable substrate for integration by Sleeping Beauty transposase. Importantly, integration with this novel vector was found to occur significantly less frequently within active genes than a standard lentiviral vector. Finally, it was shown that the transposase protein can be incorporated into lentiviral vector particles in a manner analogous to HIV-1 integrase. The development of vectors with safer integration patterns may lead to better clinical outcomes for patients treated with gene therapy.

Publications Arising

The results described in Chapters 4 and 5 of this thesis are published in:

Vink CA, Gaspar HB, Gabriel R, Schmidt M, McIvor RS, Thrasher AJ, Qasim W. Sleeping Beauty Transposition From Nonintegrating Lentivirus. Mol Ther 2009; 17(7):1197-1204. (1)

Acknowledgements

My deepest thanks go to my principal supervisor Waseem Qasim for his encouragement and support throughout the course of this project and for his helpful intellectual contributions to the writing of this thesis. I would also like to thank my supervisors Adrian Thrasher and Bobby Gaspar for their valuable guidance of the direction of research.

I am grateful to our collaborator Richard Gabriel from the laboratory of Manfred Schmidt at the German Cancer Research Centre in Heidelberg for pyrosequencing our integration site libraries. I am also indebted to João Metelo from the Institute of Child Health for his patient and selfless assistance with cell sorting and confocal microscopy. I must also thank Luís Apolonia from the Institute of Child Health for collaborating on the protein transduction research as well as his priceless intellectual contributions to the project as a whole.

On a personal note, I would like to thank Ahad for saving Albert's life after I put him in mortal danger, Albert for providing positive controls and teaching me the difference between branched and linear PEI, Anil for convincing me that there is more to life than money, Anne-Marie for keeping the heat off by being the golden girl, Aris for his advice on life and careers, Austen for showing that 29 hour workdays are possible (if you are Austen), Brian for singlehandedly keeping the lab running, Chiara Bac for saving the ladies of London from loose cannon Rahim, Chiara Be for doing some real science unlike the rest of us, Christine R for making MIU a classier place to be, Christoph for being the coolest postdoc, Claire for valuable mortgage advice, Claudia for making the windowless RV lab a nicer place to be, Dale for helping us supplement our stipends through poker and (as yet unrealised) pub quiz winnings, Dip for showing that there is life after the PhD (sort of), Eliot for not being afraid to clone while wearing a manbag, Emma and Sylvie for ignoring whiskey tongue which was Scott's fault anyway, Fang for her twin passions of opera and gossip, Gerben for making the rest of us look bad, Giorgia for keeping my boss occupied with chitchat, Helen for always wanting to go dancing, Hong for being more concerned for my safety than I ever was, Jen for being a helpful bench buddy, Jim for dragging everyone to the (old) Rugby on Friday, João M for teaching me that the Portuguese

way is the best way, João N for livening the place up, Kenth for his reliable noon hunger pangs, Luís for his friendship and wise advice over well-deserved science beers, Meng for doing the goriest experiments, Michael for being my first (and most eager) student, Michele for looking after Albert and Shahla, Mike for inspiring confidence in generations of MIU PhD students, Mohamed for keeping me on the straight and narrow, Nat for being my most zealous cycling convert, Nicky B for bringing Oxford Street to Dorset, Nicky C for her thoughts on New Labour, Roman for late night lab picnics, Scott for being the original Bad Boy, Shahla for being just about the nicest and most selfless person you could ever hope to meet, Shamika for brightening up the lab, Sneha for our good chats over peanut butter toast, Sophie for being one of the cool ones, Ste for keeping the department virus chat going against all odds, Stephanie for finding happiness, Sue for saving us all from ourselves, Suzy for being the exception that proves the rule, Tanja for her thoughts on life, Taylor for rescuing me from the desert island that is cloning, and Yin for only causing me to faint in the lab once.

I would also like to thank Luísa, from whom I stole the time to write this thesis, but who punished my crime with kindness and understanding.

For my parents:

Thanks for the DNA.

Table of Contents

List of Figures

List of Tables

Abbreviations

AAV	adeno-associated virus
ADA	adenosine deaminase
AIDS	acquired immunodeficiency syndrome
bp	base pair
BSA	bovine serum albumin
CA	capsid
CCR5	chemokine receptor 5
CD	cluster of differentiation
cDNA	complementary DNA
CGD	chronic granulomatous disease
CMV	cytomegalovirus
CpG	cytosine nucleotide adjacent to guanine nucleotide
cPPT	central polypurine tract
DAPI	4'-6-diamidino-2-phenylindole
dATP	deoxyadenosine triphosphate
DBD	DNA-binding domain
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DiI	1,1'-dioctadecyl-3,3,3'3'-tetramethylindocarbocyanine perchlorate
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleotide triphosphate
DSB	double strand DNA break
dTTP	deoxythymidine triphosphate
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescent protein
ELISA	enzyme-linked immunosorbent assay
Env	envelope
ES	embryonic stem cell
HIV-1	human immunodeficiency virus type 1
HLA	human leukocyte antigen
HR	homologous recombination
HRP	horseradish peroxidase
IL2RG	interleukin 2 receptor common gamma chain
ILV	integrating lentiviral vector
IN	integrase
IR	inverted repeat
kb	kilobase
LB	Luria-Bertani
LEDGF	lens epithelium-derived growth factor
LMO2	LIM domain only 2
LM-PCR	ligation-mediated PCR
LTR	long terminal repeat
MA	matrix

MHC	major histocompatibility complex
MLV	murine leukaemia virus
MOI	multiplicity of infection
MoMLV	Moloney murine leukaemia virus
mRNA	messenger RNA
MVB	multivesicular body
NC	nucleocapsid
Neo	neomycin phosphotransferase
NHEJ	non-homologous end joining
NILV	non-integrating lentiviral vector
NK	natural killer cell
NLS	nuclear localisation signal
nt	nucleotide
PBS	primer binding site or phosphate buffered saline
PBST	phosphate buffered saline with Tween
PCR	polymerase chain reaction
PEI	polyethylenimine
PIC	preintegration complex
PolII	RNA polymerase II
p24	HIV-1 p24 capsid protein
polyA or pA	polyadenylation
PPT	polypurine tract
PR	protease
PTD	protein transduction domain
PVDF	polyvinylidene fluoride
qPCR	quantitative PCR
R	repeat region
rAAV	recombinant AAV
RBE	Rep binding element
RNA	ribonucleic acid
RNAi	RNA interference
RNase	ribonuclease
RPMI	Roswell Park Memorial Institute medium
RRE	Rev response element
RT	reverse transcriptase
RTC	reverse transcription complex
RT-qPCR	reverse transcription quantitative PCR
SB	Sleeping Beauty
SB10 or SB11	Sleeping Beauty transposase (1st or second generation)
SCID	severe combined immunodeficiency
SCID-X1	X-linked severe combined immunodeficiency
SFFV	spleen focus-forming virus
SIN	self-inactivating
SU	surface subunit of the envelope protein
SV40	simian virus 40
TA	thymidine nucleotide adjacent to adenosine nucleotide
TAE	Tris-acetate-EDTA
T-ALL	T-cell acute lymphoblastic leukaemia
TM	transmembrane subunit of the envelope protein
tRNA	transfer RNA

TU	transducing unit
U3	unique at 3’ region
U5	unique at 5’ region
VSV-G	vesicular stomatitis virus glycoprotein
WPRE	woodchuck hepatitis virus posttranscriptional regulatory element
ZFN	zinc finger nuclease
Ψ	RNA encapsidation signal

2. Introduction

This chapter describes the intellectual background to the work presented in this thesis, beginning with a history of gene therapy and the problem of genotoxicity in clinical trials due to vector integration. This is followed by a discussion of the factors contributing to insertional mutagenesis following vector integration. The major classes of integrating vectors are described. Finally, the reasons for developing a hybrid lentivirus-transposon vector are outlined.

2.1. Gene Therapy

Gene therapy involves the introduction of genetic material into cells in order to treat or prevent disease. Classical gene therapy has been described as "using DNA as a drug", in which DNA carrying genes is transferred into cells by artificial means. After decades of research, this approach has now been successfully used to treat a number of conditions in humans. This section presents the historical background to the development of gene therapy.

Genes as theoretical units of inheritance were first described around 1900 by the botanists Erich von Tschermak, Hugo de Vries, and Carl Correns based on their observations of phenotypic segregation ratios in plants. In 1911, Thomas Hunt Morgan's studies of sex-linked traits in fruit flies led him to suggest that the genes responsible were carried on the sex chromosomes and that all genes were associated with particular chromosomes.

Although the physical location of genes was known to be chromosomes, for decades the biochemical nature of genes was controversial. Chromosomal proteins were a leading candidate owing to their complex chemical compositions. The problem was solved by Oswald Avery in 1944 when he isolated a chemical 'transforming principle' able to stably transform a strain of Pneumococcus Type II (small, rough colonies) into Pneumococcus Type III (large, smooth colonies) (2). He identified the transforming principle as "sodium desoxyribonucleate", or in modern nomenclature, DNA. Avery's work was later supported by Hershey and Chase's demonstration in 1952 that bacteria

infected by radiolabelled bacteriophage took up only the DNA of the virus, not the protein shell (3).

The structural basis by which information could be stored in DNA was suggested by the DNA double helix discovered by Franklin, Watson and Crick in 1953 (4). This discovery marked the beginning of the "Revolution in Molecular Biology" (5), a period of rapid progress in the 1950s and 60s from which emerged the Central Dogma of molecular biology – that information flows from DNA to RNA to protein – as well as the triplet genetic code described by Crick, Nirenberg, Khorana, and Brenner.

This progress in the theory of molecular biology was paralleled by advances in the ability of molecular biologists to practically manipulate DNA itself. Over the course of the 1960s, Nathans, Arber, and Smith isolated restriction endonucleases from bacteria and characterised their ability to cleave DNA in a sequence-specific manner. In 1968, Khorana used T4 DNA ligase to join DNA molecules together (6). Work by Cohen, Boyer, and Berg led to the production of the first recombinant DNA molecule in 1973, a plasmid into which novel antibiotic resistance genes were inserted using the restriction enzyme EcoRI and T4 DNA ligase (7).

The potential for the emerging technology to be applied outside of the laboratory was noted at the time. In a 1964 perspective paper, Edward Tatum declared (8):

"Within the next hundred years great advances can be expected in the control of mutational processes, in the design and synthesis of genetic determinants, and in the development of techniques for the introduction of such new genetic determinants into the genome of living organisms. The next centennial program of this Academy may very well include a symposium on "Genetic Engineering and Controlled Evolution." With the increase of our knowledge in these and other areas, it is sincerely to be hoped that our ability to use this knowledge wisely will increase in proportion."

The (at the time, theoretical) introduction of genes into human beings to treat disease came to be known as gene therapy. As Friedmann notes in his history of gene therapy (9), the term 'gene therapy' itself first appeared in the Medline index around 1970 (Figure 2.1).

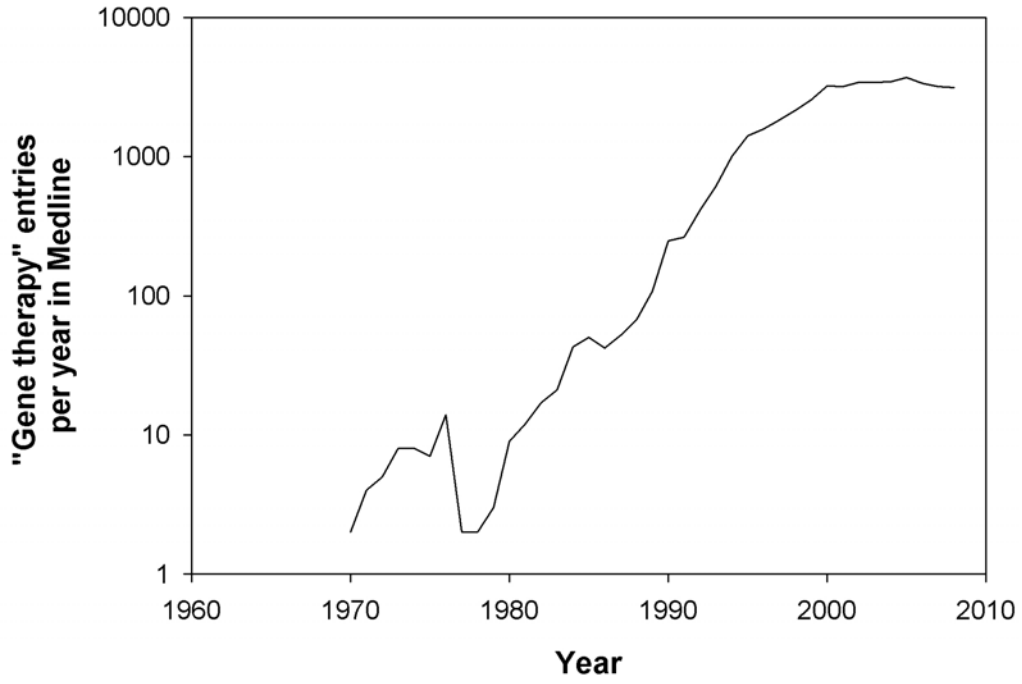

Figure 2.1 Publication trends in gene therapy.

A search of the Pubmed database was performed for each year with the search term "gene therapy".

A significant technical hurdle for the realisation of the gene therapy promise is the problem of inserting foreign DNA into mammalian cells. In 1961 Kay reported that cultured cells could take up naked DNA from the surrounding medium (10), but the process was extremely inefficient. A number of transfection reagents were later used to improve the efficiency of DNA uptake, and the first reasonably efficient solution to the problem was the calcium phosphate technique reported by Graham and coworkers in 1973 (11). In 1978, beta-globin became one of the first human genes to be cloned (12), and a number of authors were able to show that this and other functional human genes could be transfected into cells in culture (13).

It was generally thought that calcium phosphate transfection was too inefficient to be clinically useful, but in 1979 Martin Cline at UCLA was able to show stable gene transfer of a drug resistance marker into mouse bone marrow transplants (14). In 1980, Cline embarked on the first human gene therapy clinical trial in which bone marrow from beta-thalassaemia patients was transfected *ex vivo* with plasmids encoding the human beta-globin gene. Although no adverse effects were observed, no clinical benefit resulted either and the trial attracted much criticism for Cline's failure to obtain permission for human studies from his university's review board, as well as

the fact that the trial was performed in Israel and Italy, countries which had not yet established specific regulations for gene therapy trials. Cline subsequently lost his university chair and was stripped of most of his funding (reviewed in (15)). The controversy catalysed a significant public debate concerning the ethics and regulation of gene therapy trials, leading to the establishment of the National Institutes of Health Gene Therapy Subcommittee in 1984 for the regulation of gene therapy trials in the United States. The equivalent regulatory body in the United Kingdom, the Gene Therapy Advisory Committee, was established in 1993 on the advice of the Clothier Committee.

The clear need for better vectors for the introduction of transgenic DNA into mammalian cells stimulated significant progress between 1979 and 1984, a brief period from which emerged virtually all of the major classes of gene therapy vector in use today. Viruses, as naturally-occurring vehicles for the introduction of foreign DNA into cells, were seen as promising candidates for vector development. This promise was realised with the publication of the first viral vector based on Simian Virus 40 in 1979 (16) followed in rapid succession by gammaretroviral vectors in 1981 (17), adenoviral vectors in 1984 (18), and adeno-associated virus (AAV) vectors in the same year (19). A number of novel non-viral vectors were also first reported in this period, notably liposomes (20) and electroporation (21) in 1984.

These vector systems were later refined, for example by splitting viral genomes so that viral protein coding sequences could be removed from vectors (22), the development of retroviral vectors based on the lentivirus HIV-1 for transduction of non-dividing cells (23), and the synthesis of the polycation polyethylenimine (PEI) for greatly improved non-viral transfection (24). A comparison of the major classes of gene therapy vectors is given in Table 2.1.

	Gamma-retrovirus	Lentivirus	Foamy virus	Herpes virus	Adenovirus	AAV	Non-viral
Nucleic acid in vector	RNA	RNA	RNA	DNA	DNA	DNA	DNA
Packaging capacity	~9kb	~10kb	~12kb	>30kb	~30kb	4.6kb	Unlimited
Tropism	Broad	Broad	Broad	Neurotropic	Broad	Broad	Broad
Packaging cell lines	Good	Poor	Moderate	-	-	Poor	-
Integration into host genome	Yes	Yes	Yes	No	No	Rarely	Rarely
Rearrangement of transgene	+	-	-	-	-	-	-
Duration of transgene expression	Long	Long	Long	Transient	Transient	Long in postmitotic tissues	Transient
Transduction of postmitotic cells	-	+	+	+++	+++	++	++
Pre-existing host immunity	None	None	None	Yes	Yes	Yes	None
Safety concerns	Insertional mutagenesis	Insertional mutagenesis	Insertional mutagenesis	Inflammatory response	Inflammatory response	Low risk of insertional mutagenesis	-
Germline transmission	-/+	+	?	-	-	+/-	-

Table 2.1 Properties of gene therapy vector types.

Adapted from (25). AAV, adeno-associated virus; -, zero; -/+, equivocal; +, low; ++, moderate; +++, high.

The development of viral vectors improved the efficiency of gene transfer to the point where a clinical benefit from gene therapy could reasonably be expected and enabled the first fully regulated clinical trial in 1990 (26). In this trial, two children with severe combined immunodeficiency caused by a lack of adenosine deaminase activity (ADA-SCID) were treated by *ex vivo* retroviral transfer of the ADA gene into autologous T-lymphocytes which were subsequently returned to the patients. The use of autologous cells for immune transplants is preferable as it reduces the risk of acute graft versus host disease (27). No adverse effects were observed in the ADA-SCID T-lymphocyte trial, and significant expression of ADA was detected in gene-modified cells recovered from the patients. However, transient transgene expression necessitated regular infusions of gene-modified cells, and the effectiveness of gene therapy may have been masked by continued use of polyethylene glycol-ADA enzyme replacement therapy, possibly removing the selective survival advantage conferred to gene-modified T-lymphocytes relative to their unmodified counterparts (28).

The number of clinical trials initiated increased significantly over the course of the 1990s, though by the end of the decade none had reported a clear and lasting clinical benefit. In 1999 the first serious adverse event due to a gene therapy protocol occurred during a trial using an adenoviral vector to treat the liver metabolic disorder ornithine transcarbamylase deficiency. An 18-year-old man named Jesse Gelsinger suffered a massive systemic inflammatory response to the adenovirus-5 serotype vector, leading to multiple organ failure and death within days of vector administration (29).

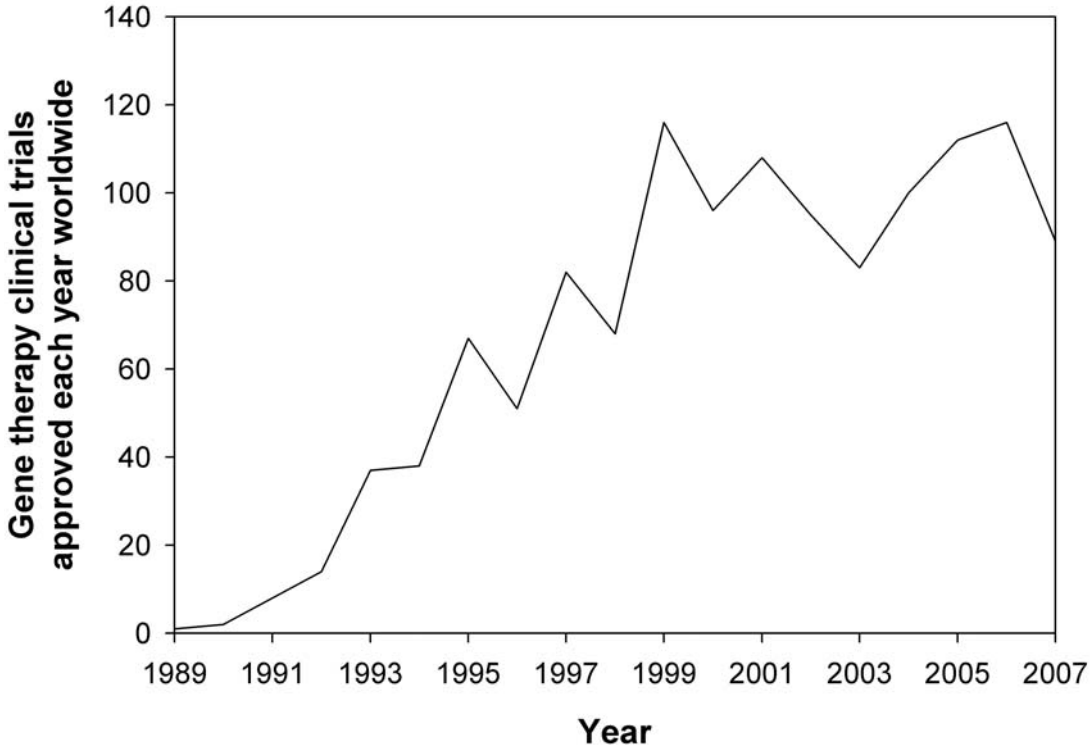

Figure 2.2 Gene therapy clinical trials approved worldwide 1989-2007.
Source: Journal of Gene Medicine Clinical Trials database (30).

The first clear success of gene therapy was reported in a French trial for X-linked severe combined immunodeficiency (SCID-X1) in 2000 (31). In this trial, the bone marrow of ten infants was extracted and cells were selected for the CD34 marker to enrich for haematopoietic progenitors. These were transduced *ex vivo* with a retroviral vector encoding the interleukin-2 receptor common gamma chain and infused into patients intravenously. In all but one patient, the protocol resulted in a normal T-lymphocyte count within months of treatment and an antigen-specific response to immunisation. The effective immune system reconstitution enabled the withdrawal of immunoglobulin therapy in the majority of patients. A similar trial to treat ten patients in Britain also resulted in effective immune system reconstitution (32). However, both trials experienced serious adverse events in the form of T-cell leukaemia-like expansions originating in the thymus (33-35). To date, this event has occurred in four patients in the French trial, resulting in one death, and one expansion has also occurred in the British trial. It is presently believed that the initiating event in these expansions is chromosomal insertion of the retroviral vector causing dysregulated

expression of nearby proto-oncogenes through the action of an enhancer contained within the vector (36). The subject of insertional mutagenesis is discussed more fully in Section 2.2.

Other notable gene therapy clinical trials (reviewed in (37;38)) include retroviral transduction of CD34 haematopoietic progenitors for the treatment of ADA-SCID (39;40), retroviral T-cell suicide gene therapy to control proliferation following T-cell and bone marrow transplant for leukaemia (41), retroviral anti-melanoma T-cell receptor immunotherapy (42), AAV-mediated neurotransmitter production for treatment of Parkinson's disease (43), and AAV expression of RPE65 in retina for treatment of Leber's congenital amaurosis (44).

Effective gene therapy for other conditions is currently hampered by a number of technical hurdles. Firstly, patients may experience an immune response to the transgene product or the gene transfer vector itself, as was observed in a clinical trial using an AAV serotype 2 vector to treat haemophilia B (45). Secondly, efficient engraftment and expansion of cell transplants modified *ex vivo* may be limited in the absence of a significant survival advantage for transduced cells, as has been observed in clinical trials for anti-HIV gene therapy (46). Thirdly, expression of the transgene may be lost following promoter silencing, as may have occurred during a clinical trial to treat chronic granulomatous disease (CGD) (47;48). Fourthly, there may be no vector presently able to efficiently transduce the target cell population, as appears to be the case with gene therapy for cystic fibrosis (49).

Although it has taken several decades for gene therapy to justify even a fraction of the early hype, the field is flushed with optimism following recent successes in clinical trials (50). Clinical application of gene therapy remains experimental, as witnessed by the fact that no gene therapy product has yet been approved for sale by the US Food and Drug Administration. Important technical challenges remain, but many are gradually yielding to a sustained research effort. Gene therapy protocols have resulted in the deaths of two patients as well as significantly improved health for dozens of others. In this sense gene therapy is maturing not into a miracle panacea, but rather a normal medical intervention.

2.2. Insertional Mutagenesis

Gene therapy for the group of haematopoietic disorders known as severe combined immunodeficiency (SCID) represents one of the most significant achievements within the field to date. SCID is a group of primary immunodeficiencies characterised by a severe defect in T-cell production and function, often accompanied by defects in B and NK cells. SCID is inherited in an X-linked or autosomal recessive fashion with an incidence of approximately 1 in 69,000 live births (51). A number of genes are known to produce a SCID phenotype, and the molecular basis of over 80% of SCID cases is known (52). The most common form (accounting for around half of all cases) is X-linked SCID (SCID-X1) resulting from molecular defects in the IL2RG gene present on the X chromosome. IL2RG encodes the common gamma chain component (γc) which is present in a number of interleukin receptors. The loss of γc function disrupts multiple interleukin signalling pathways necessary for the development of haematopoietic cells, resulting in a T(-) B(+, but dysfunctional) NK(-) phenotype. These haematopoietic defects mean that children presenting with SCID have a severely compromised adaptive immune response to infection, and if untreated most will not survive beyond the second year of life.

The immediate management of SCID includes antimicrobial drugs or immunoglobulin or enzyme replacement therapy. Long term immune reconstitution can be achieved through an allogeneic bone marrow transplant, which enables engraftment of haematopoietic stem and progenitor cells which are able to differentiate normally to produce T-cells. If a transplant can be obtained from an HLA-matched family donor, the long-term survival rates following this procedure are over 90% (53). However, survival rates are reduced in patients where a haploidentical donor is the best that is available, owing to a longer reconstitution phase and the need to use ablative chemotherapy to condition the patient for engraftment and immune reconstitution. There was previously a risk of graft-versus-host disease resulting from allogeneic T-cells present in the transplant (54), although this has been effectively addressed by the use of T-cell depletion of transplants.

Gene therapy involving the transfer of DNA encoding the IL2RG gene into patient cells for autologous transplant represents an alternative to allogeneic bone marrow

transplant for SCID-X1. SCID-X1 is a particularly attractive candidate for gene therapy for a number of reasons, including the high cost of conventional treatments and a favourable risk-benefit ratio where an HLA-matched donor is not available as well as a proposed survival advantage for IL2RG-expressing cells resulting in a large and persistent population of these cells in patients after transplant. The existence of the IL2RG survival advantage is supported by a spontaneous reversion event in a SCID-X1 patient in which partial phenotypic correction resulted from a single somatic mutation (55).

Two principal trials for gene therapy of SCID-X1 have been performed, each treating 10 patients (32;56). The conduct of these two trials was very similar. Autologous bone marrow was extracted from patients, selected for CD34 to enrich for haematopoietic progenitor cells, and transduced *ex vivo* with a gammaretroviral vector carrying the IL2RG cDNA. Cells were then infused back into patients. Both trials were highly successful, resulting in engraftment and expansion of modified cells, correction of γc signalling, and significant immune reconstitution. However, four patients in the French trial and one in the British trial experienced a serious adverse event in the form of a dysregulated expansion similar to T-cell acute lymphoblastic leukaemia (T-ALL) (57-61). One of these patients subsequently died, but the others responded well to standard anti-leukaemia chemotherapy and retained a functioning adaptive immune system after treatment. The initiating event in these leukaemic events appears to have been integration of the gammaretroviral vector into host chromosomes leading to dysregulation of nearby known T-ALL proto-oncogenes.

Vector integration into host chromosomes is necessarily a mutagenic event in that it alters the primary DNA sequence of the host. Semi-random integration patterns, as observed for retroviruses and DNA transposons, result in integration sites scattered widely throughout the genome. Insertion of DNA may affect functional elements already present at the integration site in a number of ways. Firstly, vectors may insert into and disrupt coding sequences, resulting in abnormal or prematurely terminated transcripts. Secondly, vectors may contain promoters able to read through the vector transcription termination site into neighbouring host genes. Thirdly, vectors may have enhancer activity able to dysregulate host promoters at distances up to hundreds of

kilobases in either direction from the integration site (62). Fourthly, vector insertion may disrupt other regulatory elements such as microRNA cistrons (63).

Disruption of host sequences involved in cell cycle control, such as proto-oncogenes or tumour suppressor genes, can result in cell transformation and tumourigenesis. Insertional mutagenesis with retroviral or transposon vectors has been successfully used to identify cancer genes in large screens of mice (64;65). As an aside, it is important to note that the vectors used in these screens differ from gene therapy vectors in that they have been designed for maximum disruption of nearby host elements, for example by incorporating strong outward-facing promoters and splice donor or acceptor sites. Nonetheless, integration of conventional retroviral vectors resulting in transformation has been observed in both animal models (66) and two gene therapy clinical trials (67;68). Understanding and preventing this process is one of the major challenges facing the gene therapy field at the present time.

Although the precise mechanism of leukaemic transformation in the SCID-X1 trials continues to be the subject of much study, a great deal is already known. In all cases, a latent period on the order of years was observed between transplantation of gene modified cells and subsequent leukaemic expansions. At the time of expansion, the polyclonal T-cell population became dominated by one or a small number of T-cell clones. In four of the five leukaemic patients, dominant clones were found to contain retroviral insertions within or near the known T-ALL proto-oncogene LMO2 (69;70). This gene was overexpressed in mature T lymphocytes, probably as a result of the enhancer activity of the vector promoter. Insertions near other T-ALL proto-oncogenes such as LYL1 have also been identified. A general model has been proposed in which insertional mutagenesis leads to continued expression of developmental genes which are normally expressed in haematopoietic stem cells but subsequently downregulated during immature T-cell development, and this continued expression disrupts the normal T-cell expansion and maturation processes in the thymus (71). Although a single retroviral insertion appears to be sufficient to initiate progression to leukaemia, other insertions and acquired somatic mutations most likely contributed to the observed expansions (72).

Interestingly, clonal dominance in granulocytes caused by retroviral integration near proto-oncogenes was also observed in a gene therapy clinical trial to treat CGD,

though in this case it is believed the expansions contributed to the temporary therapeutic benefit observed in these patients by increasing the number of gene corrected cells in circulation (73).

Much effort has been made to understand the risk factors underlying these events so that the lessons can be applied to future gene therapy trials for both SCID-X1 and other conditions. The major suspects in SCID-X1 leukaemogenesis include the immune deficiency of the patients, haematopoietic 'stress' experienced during immune reconstitution, the enhancer activity of the vector promoter, the proposed oncogenic potential of the IL2RG transgene, and the integration preferences of gammaretroviral vectors.

It is known that immunodeficiency increases the risk of malignancy, and it has been suggested that the low NK cell number following SCID-X1 gene therapy might have compromised immunosurveillance of leukaemic cells (74;75). However, an increased incidence of leukaemogenesis has not been reported in patients receiving an allogeneic bone marrow transplant for SCID-X1. Typically, these patients also have low NK cell counts.

It has been suggested that transplanted cells experience haematopoietic 'stress' due to rapid expansion and forced clonal selection for gene modified cells (76). For example, murine bone marrow transduced with retroviral vectors expressing a reporter gene can become leukaemic following serial transplantation (77). A rhesus macaque which received a bone marrow transplant marked with a drug resistance gene developed leukaemia some years after cytotoxic chemotherapy to select for gene modified cells (78). The dramatic selective proliferation advantage conferred to engrafted cells during SCID-X1 gene therapy may have resulted in a similar clonal imbalance.

The best-defined risk factor in the SCID-X1 leukaemic events remains overexpression of the LMO2 proto-oncogene (79), most likely driven by the enhancer activity of the gammaretroviral vector long terminal repeat (LTR) promoter. It seems probable that this event is particular to the biology of SCID-X1, though other tissues and conditions may have their own oncogenic soft spot as was shown by a common insertion site present in mouse hepatocellular carcinomas following insertional mutagenesis by an AAV vector (80). Modifications to the vector design which minimise enhancer

activity may improve the safety of these vectors. For example, self-inactivating gammaretroviral vectors in which the strong LTR promoters are replaced with an internal promoter show reduced transactivation of neighbouring genes and *in vitro* immortalisation, though a lower level remains if the internal promoter also possesses enhancer activity (81). Reduction of enhancer activity may be achieved through careful choice of the internal promoter (82), and also by placing chromosomal insulators within vectors in order to prevent enhancers within the vector from acting on neighbouring host genes (83;84).

Constitutive retroviral expression of the IL2RG transgene has been argued to be oncogenic. A mutagenesis screen identified a tumour carrying insertions at both IL2RG and LMO2 (85), and a study was performed in which SCID-X1 mice treated with a lentiviral vector expressing IL2RG developed T-cell lymphomas while the enhanced green fluorescent protein (eGFP) expressing control mice did not (86). This study has been disputed by others on the grounds that the high vector copy numbers and overexpression of IL2RG in this model did not accurately reflect the conditions of the clinical trials (87).

Lastly, the integration site preferences of the gammaretroviral vectors used in the SCID-X1 trials may have contributed to their leukaemic potential. It is known that gammaretroviral vectors integrate preferentially near transcription start sites and CpG islands (88). The level of risk due to these preferences relative to randomly distributed integration sites is not known, but alternative integrating vectors with different preferences do exist. For example, lentiviral vectors integrate preferentially within active genes (89) and the Sleeping Beauty transposon has an almost random integration profile (90). A direct *in vitro* or *in vivo* comparison of the oncogenic potential of these different integration patterns is difficult because these vectors differ greatly in their target cell tropism. In a tumour-prone mouse model, self-inactivating lentiviral vectors carrying an enhancerless internal promoter showed no detectable increase in tumour formation, in contrast to gammaretroviral vectors carrying full LTR promoters (91). This study did not address the relative roles of the promoters and integration profiles in tumourigenesis. Rarely, patients infected with wildtype HIV develop lymphomas associated with a common chromosomal insertion site (92), but the relevance of this observation for pseudotyped lentiviral vectors in other cell types

is not clear. An understanding of the role played by integration site preferences in vector oncogenic potential requires a like-for-like comparison in relevant *in vivo* models for insertional mutagenesis, and such studies have not yet been reported.

2.3. Integrating Vectors for Gene Therapy

A number of vectors currently being developed for gene therapy actively integrate the vector DNA into target cell chromosomes. The main advantage of integration is that it is capable of yielding stable, long-term transgene expression, particularly in mitotic tissues.

All episomal nuclear DNA is probably capable of 'background' integration into chromosomes under the action of host DNA repair proteins. For example, adenoviral vectors integrate into chromosomes at low frequency ($\sim 10^{-4}$ integrations per intracellular vector genome), often through illegitimate recombination between regions of microhomology on both vector and chromosome (93). However, this phenomenon is likely to be too inefficient to be useful for most applications. By contrast, integrating vectors actively catalyse integration using vector-derived proteins. This section describes the major classes of integrating vector employed to date.

2.3.1. Retroviral Vectors

The family of retroviruses is known as *Retroviridae* and consists of a number of enveloped positive-sense single-stranded RNA viruses for whom reverse transcription and chromosomal integration of the viral genome are essential stages of the cell cycle. Within this family, the gammaretrovirus, lentivirus, and spumavirus (foamy virus) genera have been developed as vectors for gene therapy.

The retroviral vector used in this study was an HIV-1-based lentiviral vector, so the following summary is focused principally on this virus. However, where important differences exist between lentiviruses and gammaretroviruses or foamy viruses, these are described.

2.3.1.1. The Molecular Biology of HIV-1

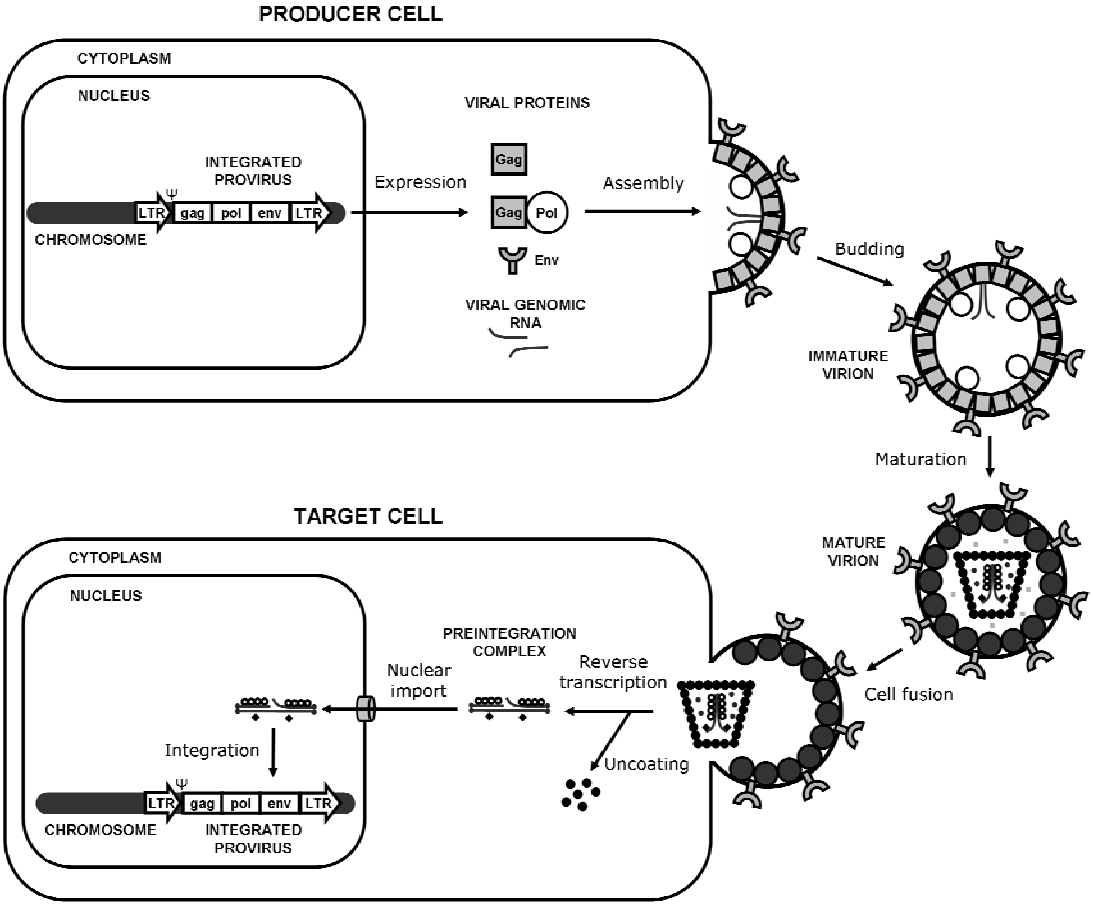

Figure 2.3 Life cycle of a simple retrovirus.

Over the course of their life cycle, retroviruses alternate between two major forms known as the provirus and the virion (Figure 2.3; reviewed in (94)). The provirus consists of double-stranded DNA integrated into a host cell chromosome. Viral RNA and proteins are expressed from the provirus using the host's own transcription and translation apparatus. These are subsequently packaged at the host plasma membrane into virion particles which are enveloped by a host-derived lipid membrane. The resulting virion can bind to and enter a new host cell, reverse transcribe its genome to regenerate the DNA form, and then integrate this into a host cell chromosome as a new provirus.

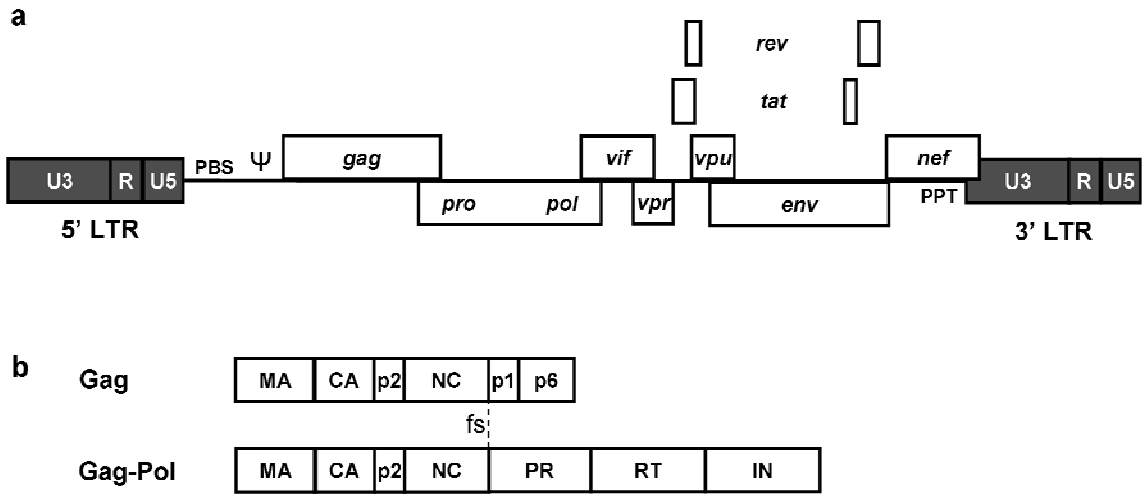

Figure 2.4 HIV-1 provirus and polyprotein structure.

(a) Structure of an HIV-1 provirus. LTR, long terminal repeat (subdivided into U3, R, and U5 regions); PBS, tRNA primer binding site; Ψ, RNA packaging signal; gag, polyprotein encoding virion structural components; pro-pol, polyprotein encoding viral enzymes; vif, vpr, vpu, and nef, accessory genes; rev and tat, regulatory genes; PPT, polyprurine tract. **(b)** HIV-1 Gag and Gag-Pol polyproteins. MA, matrix; CA, capsid; p1, p2, and p6, spacer peptides; NC, nucleocapsid; fs, ribosomal frameshift site; PR, protease; RT, reverse transcriptase; IN, integrase.

Transcription

Transcription of proviral DNA by the host RNA polymerase II (PolII) enzyme is driven by the U3 promoter within the 5' long terminal repeat (LTR). Most of the regulation of transcription occurs through host transcription factors acting on binding sites present within the U3 promoter (compiled in (95)). HIV-1 regulates its own transcription through the virus-encoded Tat protein. In the absence of the Tat protein, viral transcription from the HIV-1 5' LTR yields short, non-polyadenylated RNA (96;97). Tat binds to a *trans*-activator response region which forms a stem-loop structure within the 5' R region of the nascent viral RNA (98). It is thought that this interaction allows Tat to recruit host proteins concerned with transcription and chromatin remodelling, allowing viral transcription to continue normally. Gammaretroviruses do not regulate their transcription in this way, but foamy viruses express a transactivator protein with a similar function (99).

During processing the transcript receives a 5' cap (100) and a 3' poly-A tail (101) using the same machinery as host PolII-transcribed mRNA. HIV-1 produces at least

thirty different mRNAs through alternative splicing at approximately four splice donor and eight splice acceptor sites, though at least half of its mRNA remains entirely unspliced (102). This complexity enables HIV-1 to express multiple genes from a single promoter, and also contributes significantly to the plasticity of HIV-1 gene expression. Splicing in gammaretroviruses is less complex, resulting in only two mRNA species.

Splicing of cellular mRNAs is linked to their nuclear export. Typically, unspliced host mRNAs remain in the nucleus until they are degraded. The unspliced virus mRNA species is necessary for both the production of the proteins it encodes and its role as the virion genome, so specialised nuclear export mechanisms have evolved. The Rev protein of HIV-1 binds to the Rev response element (RRE) on unspliced viral RNA, mediating its nuclear export (103) via a leucine-rich nuclear export signal (104). Gammaretroviruses do not express a Rev protein, but the packaging signal (Ψ) may contribute to nuclear export of unspliced mRNA (105).

Translation

Translation of retroviral mRNA is carried out by host ribosomes. The HIV-1 Rev protein has been shown to enhance the translation of RRE-containing mRNA through increased ribosomal association (106). This may be a means of compensating for the low basal rate of translation due to the very different codon usage in HIV-1 and humans (107).

Although viral transcripts carry a 5' cap, initiation of translation does not occur through the normal ribosome-scanning mechanism, possibly because there is significant RNA secondary structure upstream of the translation start site. Instead, an internal ribosome entry site (IRES) upstream of *gag* may be used to initiate translation of unspliced transcripts (108).

Translation of unspliced viral mRNA gives rise to the Gag and Gag-Pol polyproteins. Gag contains the structural proteins matrix (MA), capsid (CA), and nucleocapsid (NC). In addition to these, Gag-Pol contains the enzymes protease (PR), integrase (IN), and reverse transcriptase (RT). Gag and Gag-Pol are translated from the same mRNA transcripts in a ratio of approximately 20:1 (109). Translation of Gag-Pol

requires bypassing of the Gag termination codon. HIV-1 achieves this through ribosomal frameshifting, whereby the ribosome slips back one nucleotide into the Pro-Pol reading frame when it reaches the end of *gag* (110). This process is enabled by the presence of a 7 nucleotide 'slippery' sequence which allows the slippage and subsequent restoration of base-pairing between the A and P site tRNAs and the transcript (111) and is augmented by an RNA stem loop downstream of the shift site (112).

Gag is myristoylated during translation in the cytosol (113). This involves the attachment of the 14-carbon fatty acid myristate to the N-terminal glycine of Gag. The myristate moiety acts together with a basic N-terminal membrane binding (M) domain (114) to allow Gag to become membrane-associated.

The retroviral envelope protein is encoded by *env*. This is translated as a polyprotein from a spliced mRNA. Like virtually all host transmembrane proteins, the Env polyprotein carries an N-terminal signal peptide which allows targeting to the rough endoplasmic reticulum by the signal recognition peptide. The C-terminal region contains a stretch of hydrophobic amino acids that are inserted into the membrane, where they act as a transmembrane anchor (115). Env is glycosylated in the endoplasmic reticulum during translation. This occurs at 24 sites in HIV-1 (116), and the sugars are subsequently modified by host enzymes. Glycosylation moieties are critical for correct Env folding and cleavage but Env remains functional if they are removed after these processes are complete (117;118). In addition to its role in Env processing, glycosylation probably acts to protect the peptide backbone of the envelope protein from neutralising antibody recognition (reviewed in (119)).

Env oligomerises in the Golgi apparatus, thus becoming competent for receptor binding. Recent results suggest that the HIV-1 Env polyprotein forms trimers (120). In their oligomeric form, the polyproteins are cleaved by host proteases at the furin recognition site (121), resulting in heterodimers of surface (SU) and transmembrane (TM) proteins. This cleavage exposes the fusogenic peptide of TM (122), making the Env oligomers competent for cell fusion.

All retroviruses express the basic genes *gag, pro, pol, and env*. In addition, HIV-1 expresses the essential regulatory genes *tat* and *rev* described above and the four

accessory genes *vif, vpr, vpu,* and *nef.* The accessory genes are often multifunctional and their diverse roles in the HIV-1 life cycle are an area of active research (reviewed in (123)). However, a unifying theme that has recently emerged is that HIV-1 accessory genes frequently act to protect the virus from restriction factors, host proteins which serve as antiviral defences at multiple points in the viral life cycle.

Vif is thought to be a viral countermeasure against the antiviral activity of the host cytidine deaminase APOBEC3G (124). In the absence of Vif, APOBEC3G is packaged into virions (125), where it reduces virion infectivity by editing the viral genome (126). When Vif is expressed in producer cells, APOBEC3G is targeted for degradation by proteasomes (127).

The Vpr protein has at least five proposed functions: enhancing the fidelity of reverse transcription, possibly by recruiting a host uracil DNA glycosylase to virions (128;129); assisting nuclear import of the preintegration complex (130); causing cell cycle arrest at the G2/M boundary (131;132); activating transcription from the HIV-1 LTR promoter (133); and cell killing (134;135).

Vpu may protect the provirus from superinfection by targeting the viral receptor CD4 for degradation (136;137) and also enhances the release of virions from cells by counteracting the host restriction factor tetherin (138).

Among other functions, Nef also reduces CD4 presentation at the cell surface (139). It downregulates the cell surface presentation of MHC class I (140) and class II (141). The advantage of these activities to the virus is not well understood. Nef is also thought to assist infectivity by enhancing penetration of the viral core through the actin network adjacent to the cell surface (142).

Virion Assembly and Budding

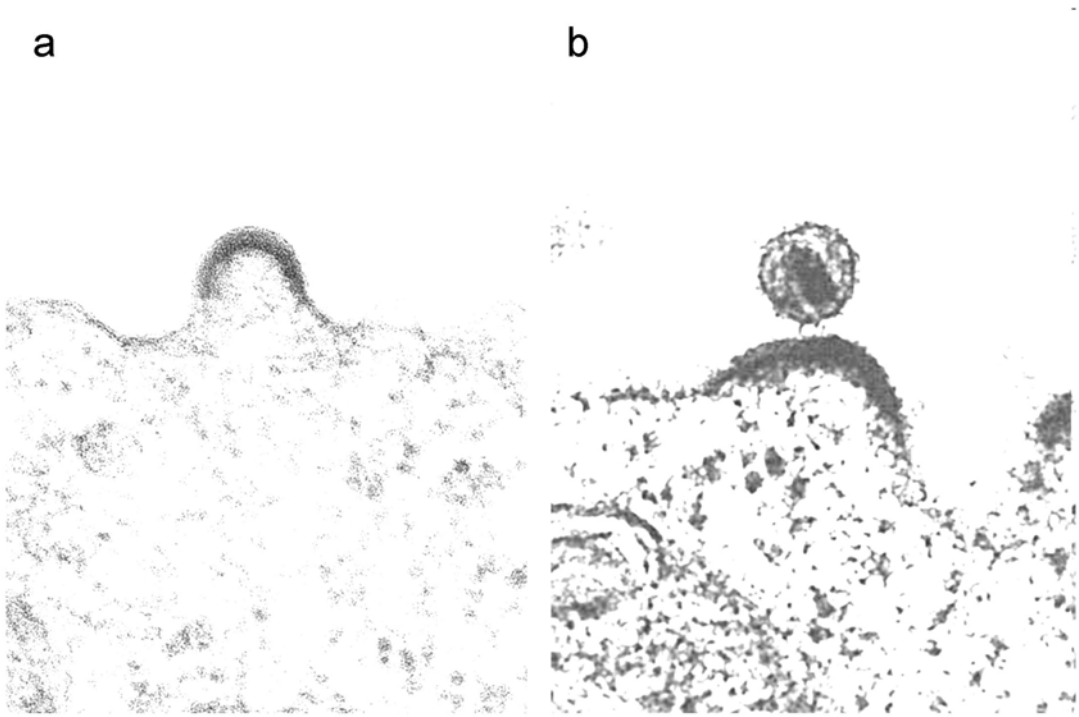

Figure 2.5 HIV-1 virion budding.
Transmission electron micrographs showing HIV-1 virions **(a)** during and **(b)** after budding from a T-lymphocyte. Source: Public domain images from (a) the Centers for Disease Control Public Health Image Library and (b) the National Institutes of Health AIDS History Image Archive.

Virion assembly requires the subcellular co-localisation of the viral Gag, Gag-Pol, and Env proteins together with the viral RNA genome and a number of essential host factors.

The Gag polyprotein is the major structural component of the immature virion and can induce the assembly and budding of virus-like particles in the absence of other viral components (143;144). Gag multimerisation is an essential step in the formation of virions. Gag-Gag interactions are mediated by the interaction (I) domain (145) and occur almost exclusively on cellular membranes following myristoylation. Gag-Pol may multimerise in the same way using its Gag region, but it can package into virions independently of this (146). There are estimated to be 2000-5000 molecules of Gag in each HIV-1 virion (147).

A number of HIV-1 accessory genes are packaged into virions. Vpr appears to be packaged in a molar ratio of about 1:7 with Gag (148). Packaging of Vpr occurs through an interaction with the p6 region of Gag (149). Nef is estimated to be present at 10 to 100 particles per virion (150). Vif is present at detectable levels in immature virions but is largely absent following maturation (151).

Unspliced viral genomic RNA is recruited to nascent virions as a dimer through an interaction between the nucleocapsid (NC) region of Gag and a packaging signal (Ψ) which for HIV-1 consists of four stem loops present in the RNA secondary structure (152).

Host factors recruited to virions include the tRNA primer for first strand DNA synthesis during reverse transcription ($tRNA^{lysine-3}$ for HIV-1) which is packaged via an interaction with the Gag-Pol polyprotein (153). Many host proteins are also packaged into retroviral virions (reviewed in (154)). Well-studied examples include HLA-II, cyclophilin A, and actin. For most of these proteins it is not known whether they serve some function for the virus, represent a host antiviral defence, or are packaged unintentionally because of their subcellular location.

Recruitment of Env to nascent virions is poorly understood. No region of Env that is necessary or sufficient for efficient packaging into virions has been defined. Paradoxically, nascent retroviral virions are able to incorporate envelope proteins from other enveloped viruses which show no homology to their own Env protein, a phenomenon known as pseudotyping (155). This suggests that rather than specific recognition between the Gag and Env proteins, co-localisation of these factors occurs within a specific subcellular structure, possibly lipid rafts (reviewed in (156)). The lipid raft model of virion assembly suggests that lipid rafts provide foci within cell membranes in which viral components interact weakly but cooperatively until they form aggregates capable of budding.

Budding is the process by which rafts of multimerised Gag and other components form into spherical bodies surrounded by a host-derived lipid membrane which has been separated from the producer cell. It has been suggested that retroviruses make use of a pre-existing host exosomal pathway (multivesicular budding, MVB) in both virion budding and cell entry (157). Exosomes are small membrane-bound bodies

which bud away from the host cell during normal cellular exchanges with the external environment. Retroviral virions are a similar size to exosomes (~100nm), and the MVB protein Tsg101 plays an essential role in HIV-1 budding through an interaction with the p6 region of Gag (158). Exosomes can form either immediately in endosomal domains at the plasma membrane or via budding into endosomes within the cell. This parallels the observation that HIV-1 buds into endosomes in macrophages and at the cell surface in T-cells (159).

Virion Maturation

During or soon after budding, virions undergo a dramatic morphological transition known as maturation. Maturation causes a shift in virion electron density from the envelope into the core. This process is dependent upon processing of viral Gag and Gag-Pol by the viral protease to yield active virus proteins. Virions are non-infectious prior to maturation.

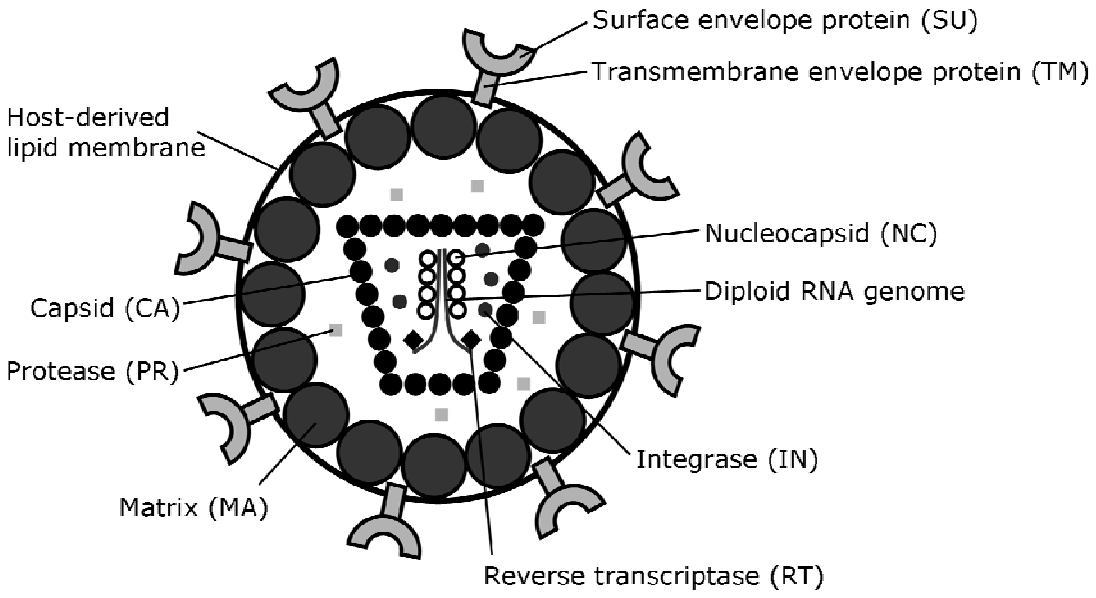

Figure 2.6 Schematic view of a mature retroviral virion.
Adapted from (160).

Processing of the polyproteins takes place in an ordered fashion due to differences in the rate of cleavage at each site. In HIV-1 Gag, cleavage occurs first between the p2 peptide and NC, then at the MA-CA and p1-p6 boundaries, and finally at the CA-p2 and NC-p1 boundaries. In Gag-Pol, p2-NC cleavage is followed by cleavage at the

RT-IN and MA-CA boundaries, and lastly sites flanking PR (161). It is not known whether the ordering of cleavage events has any biological significance. Nef is also cleaved by the viral protease during maturation (162).

In the mature virion (Figure 2.6), MA is proposed to be bound to the envelope, CA forms the outer shell of the virion core, and NC associates with the viral genome. Thus, the shift in electron density during maturation is likely to be the result of the release of CA and NC from membrane-bound Gag followed by their migration and reassembly at the virion core.

NC has been shown to promote proper annealing of the tRNA primer to the genomic RNA primer binding site (PBS) during maturation (163), probably by helping to open the RNA secondary structure (164). The mechanism of annealing is controversial (reviewed in (165)).

Cell Entry

The major determinant of viral tropism is recognition of a host cell surface receptor by the virion Env protein. However, this is not the initial contact between the virion and the cell. A number of non-receptor molecules have been implicated in initial binding of HIV-1 to the cell surface, including nucleolin (166), LFA-1 (167), heparan sulphate proteoglycan (168), and a dendritic cell surface protein known as DC-SIGN (169), perhaps allowing HIV-1 to use dendritic cells as a vehicle for transport into lymph glands in order to reach the host T cell population.

Following the initial contact, a specific interaction between Env and the host receptor is necessary for successful viral entry. The primary receptor for HIV-1 is CD4 (170), but successful infection also requires the coreceptors CCR5 (171-175) and/or CXCR4 (176), which are both involved in chemokine signalling. Initial binding between the HIV-1 Env protein and CD4 alone does not enable an interaction between the fusion peptide and the cell membrane. Instead, it is thought that the initial contact between SU and CD4 induces a conformational change in SU, revealing a strongly conserved, high-affinity coreceptor binding domain. In this way, HIV-1 is able to keep its conserved binding site buried until it is immediately adjacent to the cell surface, by which point neutralising antibody recognition is sterically impossible (177).

Coreceptor binding leads to exposure of the TM fusion peptide (178), which may allow a hydrophobic stretch of TM known as the fusion peptide to interact with the cell membrane, possibly in a manner resembling the 'spring-loaded' mechanism of influenza haemagglutinin (179).

For the viral capsid to enter the cytosol, fusion between the host and virus membranes must take place. Retroviruses employ the type I fusion pathway (reviewed in (180)). After receptor recognition, the viral TM is believed to insert its fusion peptide into the host cell membrane. It subsequently collapses into a six helical bundle (6HB) conformation, releasing free energy and bringing the membranes into close proximity for fusion (181). This initially results in a small, labile pore but later develops into a larger, robust pore that can pass the capsid into the cytosol.

As with budding, the site of virus membrane fusion is open to question. Enveloped viruses are able to fuse with host membranes at the cell surface or at endosomal membranes following endocytosis. Endocytosis can take place via clathrin-coated pits, caveolae, and pathways independent of both clathrin and caveolae such as macropinocytosis. It was initially thought that uptake of HIV-1 particles by macropinocytosis was possible but unlikely to lead to productive infection as virions were ultimately degraded in the lysosome (182). However, virus infectivity by this pathway is dependent upon the culture conditions employed (183). Associations between HIV-1 infection and clathrin-mediated (184) and caveolar (185) endocytosis have been identified.

Reduction of host membrane glycosphingolipid (186) and cholesterol (187) content is known to reduce HIV-1 fusion, suggesting a role for lipid rafts in this process. CD4 and CCR5 appear to be associated with lipid rafts during productive HIV-1 infection (188), though this association may not be necessary for successful viral entry (189). Expression of Nef in the producer cell may increase HIV-1 entry via an endocytic pathway (190), possibly by enhancing the cholesterol content of the virions (191) and its influence on fusion. However, other studies have cast doubt on the role of Nef in fusion (192;193), and Nef is not required for HIV-1 pseudotypes which enter cells via pH-dependent endocytosis (194). These seemingly contradictory results may reflect the existence of multiple pathways for cell entry of HIV-1.

Uncoating

Uncoating refers to post-entry changes in the protein composition of the intact viral core as it becomes first a reverse transcription complex (RTC) in which viral DNA synthesis occurs, and then a preintegration complex (PIC) which is competent for integration into the host genome. Uncoating, reverse transcription, and transport to the nucleus appear to be related processes. For example, intact cytoskeletal actin microfilaments are essential for efficient reverse transcription (195). However, the timing, location, and protein requirements of the post-entry processes are poorly understood.

HIV-1 cores appear to carry out most uncoating soon after entry. RT, IN, NC, phosphorylated MA (196), and Vpr have been detected in the HIV-1 RTCs and PICs, but CA appears to rapidly dissociate (197-199). Nef has been ascribed a possible role in uncoating or reverse transcription (200) and may contribute to the progression of infaction by disrupting the cortical actin network and allowing the viral core to proceed into the cell (201). In human cells, post-entry association of viral cores with target cell-derived cyclophilin A appears to promote infection. The mechanism is not known, but may be related to the action of the host TRIM5α restriction factor which acts during uncoating (202;203).

RTC/PICs are too large to diffuse through the cytoplasm at the speeds necessary for productive infection, so it is thought that they migrate towards the nucleus through interactions with the host cytoskeleton. An interaction between HIV-1 MA and actin microfilaments has been identified soon after viral entry (204), IN is known to interact with yeast microtubule associated proteins (205), and NC enables packaging of actin into budding virions (206). McDonald *et al* have visualised the movement of eGFP-tagged PICs along microtubules towards the centrosome (207). These observations are consistent with a model in which HIV-1 initially uses actin microfilaments in order to access the microtubule network.

Reverse Transcription

The requirement for reverse transcription during the replicative cycle is the defining feature of retroviruses. Most (but not all) HIV-1 RTCs are thought to complete

reverse transcription prior to nuclear entry (208). The essential catalytic activities for this process are provided by reverse transcriptase (RT). Interestingly, foamy viruses are able to reverse transcribe their DNA prior to budding from the producer cell (209).

HIV-1 RT has two major activities. Its C-terminal portion has RNase H activity which cleaves RNA, but only in RNA:DNA duplex form. The N-terminal portion carries out RNA- or DNA-dependent DNA polymerisation. The viral genomic RNA is cleaved into 6-14nt fragments by RNase H activity during minus strand DNA synthesis, but these activities are functionally uncoupled and so need not be catalysed by the same enzyme (210-212). DNA synthesis by RT is thought to be relatively error-prone, with estimated mutation rates of 10^{-4} to 10^{-5} mutations per base-pair per cycle (reviewed in (213)).

Reverse transcription begins with synthesis of minus strand DNA by RT. The primer for minus strand DNA synthesis during reverse transcription is a host-derived $tRNA^{lysine-3}$ primer (214) which is annealed to the primer binding site (PBS) through the action of NC during maturation (215). The initiation complex is initially unstable resulting in slow synthesis of DNA, but the rate of polymerisation increases 3400-fold upon addition of the seventh dNTP (216).

Minus strand DNA synthesis proceeds to the 5' end of the viral genomic RNA and is followed by transfer of the minus-strand cDNA to the RNA 3' end (217). The minus-strand cDNA product synthesised before minus-strand transfer is also known as minus-strand strong stop DNA (-sssDNA) from *in vitro* experiments, but the difficulty of detecting it *in vivo* implies that minus-strand transfer is relatively efficient (218). This transfer can occur either intra- or intermolecularly (219) and may be enhanced by the ability of NC to act as a chaperone for processes which require melting and reannealing of nucleic acids (220).

Minus-strand transfer occurs through interactions between the cDNA and 3' RNA R sequences. There is some debate as to the relative importance of R sequence homology, *cis* elements, and local secondary structure for efficient transfer (221-223). An acceptor invasion mechanism for minus-strand transfer has been proposed (224) in which initial binding between the cDNA and the 3' RNA is followed by displacement

of cleaved RNA remnants from the cDNA by branch migration. Ultimately, the cDNA 3' terminus is transferred onto the 3' RNA.

The transferred minus-strand DNA acts as a primer for continued DNA synthesis up to the 5' end of the remaining RNA (the 5' end of the PBS). As it proceeds, this process creates an RNA:DNA duplex which is cleaved by the RNase H activity of RT in a relatively sequence-independent manner. However, a polypurine tract (PPT) immediately upstream of their 3' LTR is recognised and cleaved specifically at its 3' terminus and also at or near the 5' terminus by the RNase H activity of RT (reviewed in (225)). The resultant DNA-duplexed RNA acts as the primer for plus-strand DNA synthesis by RT.

RT commences plus-strand DNA synthesis using the PPT as a primer and the minus-strand DNA as a template. The RNA-DNA junction at the 3' end of the PPT is later cleaved by a second RT enzyme (226). Plus-strand synthesis continues until it reaches the end of the tRNA PBS. The formation of a short tRNA:DNA duplex allows removal of the tRNA primer by RNase H activity (227). Following the completion of plus-strand synthesis, plus-strand transfer takes place.

The DNA strand formed from the first period of plus-strand synthesis is known as plus-strand strong stop DNA (+sssDNA). Plus-strand transfer occurs through base-pairing of the PBS on the two DNA strands (228).

Following plus-strand transfer, DNA synthesis continues in both directions to the end of the LTRs. Unlike other retroviruses, a second, central polypurine tract (cPPT) has been identified in a number of lentiviruses including HIV-1 (229). This acts as a second primer for plus-strand synthesis. After plus-strand transfer, continued synthesis through the cPPT leads to strand displacement and the formation of a 99 nucleotide single-stranded DNA flap (230). This flap was thought to play a critical role in the ability of lentiviruses to infect nondividing cells, but this has been questioned in later reports.

Nuclear Entry

Lentiviruses such as HIV-1 can integrate into and productively infect non-dividing cells (231), while gammaretroviruses such as the Moloney murine leukaemia virus (MoMLV) cannot (232). This observation led to the hypothesis that gammaretroviruses are unable to cross the nuclear envelope and so can only gain access to chromosomes for integration when the envelope breaks down during mitosis, while lentiviruses possess at least one element which permits transport of the PIC across the nuclear membrane.

The cPPT, MA, IN, and Vpr were all suggested as possible candidates, but further research has cast doubt on all of these (reviewed in (233)). The cPPT-derived central DNA flap been shown to enhance lentiviral vector-based gene transfer to non-dividing cells under some conditions (234), but its effectiveness has proven to be highly variable (235;236). MA carries two basic nuclear localisation signals (NLS) (237) which allow it to enter via the nuclear pore through interactions with importin proteins, but viruses severely depleted in MA are still able to enter the nucleus, so it does not appear to be essential for this process (238). A putative IN NLS was found (239) but later disputed (240;241). Vpr-deficient virus can productively infect nondividing cells with reduced efficiency (242).

Yamashita and Emerman constructed an HIV-1-based virus in which Vpr was removed, the cPPT was inactivated by mutation, and the HIV-1 IN and MA were replaced with their MoMLV counterparts (243). This virus showed no difference in infectivity between dividing and nondividing cells, implying that the ability of lentiviruses (but not gammaretroviruses) to infect nondividing cells has little to do with any of these proposed nuclear-localising factors. The authors argued instead that CA is the key determinant, and HIV-1 particles carrying the MoMLV CA protein are indeed unable to infect nondividing cells (244). CA is thought to dissociate relatively rapidly from HIV-1 cores during uncoating but remain associated with MLV cores up to the nuclear envelope. Complete retroviral capsids may be too large to enter the nucleus through nuclear pores.

Integration

Integration of the viral DNA into a host chromosome regenerates the provirus and is an essential step in the retroviral life cycle (reviewed in (245)). Integration is

catalysed by the integrase, a ~290 amino acid protein consisting of three domains: an N-terminal domain containing an HHCC zinc finger motif (246); the catalytic core domain in which a DDE motif coordinates the Mg^{2+} ions that catalyse integration (247); and a C-terminal domain which contains an SH3 fold involved in sequence-nonspecific DNA binding (248). It is likely that all three domains contribute to DNA binding (249;250).

Linear viral DNA in the nucleus may be converted to 1-LTR or 2-LTR circles by homologous recombination or non-homologous end joining in the nucleus. Unlike the linear form, these are not efficient substrates for the integration reaction (251;252).

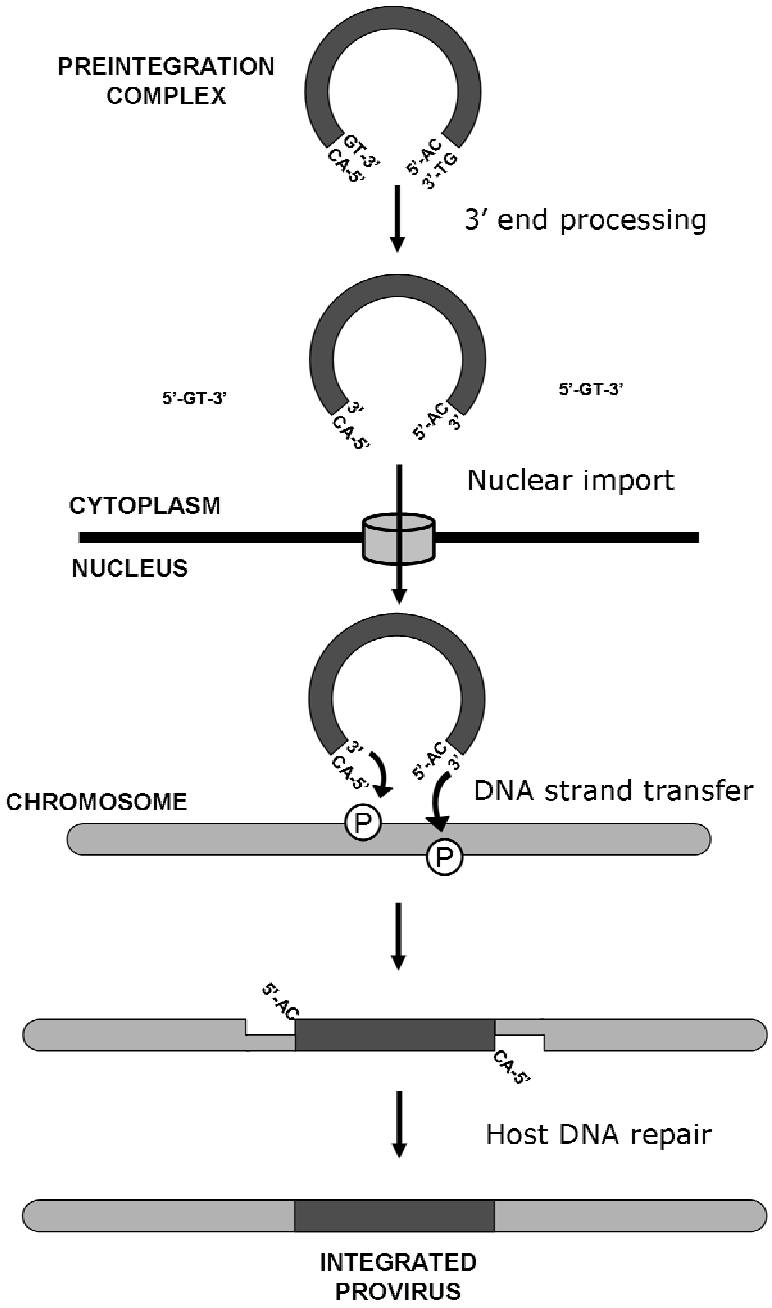

Figure 2.7 HIV-1 integration.

The 3' ends of the preintegration complex are first processed by HIV integrase, releasing two GT dinucleotides. The resulting recessed 3' hydroxyl groups attack phosphate groups on the target chromosome during the strand transfer reaction. Host DNA repair proteins subsequently remove the dinucleotide overhangs and fill in the single stranded DNA, resulting in a short duplication of host sequence flanking the integrated provirus. P, phosphate group in DNA backbone. Adapted from (253).

Two nucleophilic attacks comprise the integration reaction mechanism at each end of the linear DNA molecule. Initially, the Mg^{2+} ion of integrase deprotonates water molecules, activating them as nucleophiles able to attack near each 3' end of the viral

DNA, displacing two nucleotides and exposing a conserved CA dinucleotide. This reaction is known as 3' end processing (254). It is thought to take place very soon after the completion of reverse transcription, which for most PICs means prior to nuclear entry. Subsequent nuclear entry and binding of integrase to a host cell chromosome brings the viral and target DNAs into close proximity. The 3' hydroxyl group on the processed viral LTR is sufficiently nucleophilic to attack the target DNA strand. This is known as the strand transfer reaction. Both LTRs attack the target DNA at sites a few base pairs apart, resulting in a joining intermediate (255). The intermediate is resolved by unpairing the target strand bases between the two positions at which the strands are joined. This results in two sections of complementary, unpaired sequence at either end of the proviral insertion which are most likely repaired by host DNA repair factors (256), producing a 5bp duplication of host sequence at either end of the insertion.

A number of host proteins may participate in HIV-1 integration (reviewed in (257)). Significant among these is the chromatin-tethering factor LEDGF/p75 (reviewed in (258)). In addition to its tight association with chromatin throughout the cell cycle, LEDGF/p75 binds to HIV-1 integrase via a C-terminal integrase binding domain (IBD) (259). Knockdown of LEDGF/p75 expression or overexpression of dominant negative LEDGF/p75 containing an IBD but no chromatin binding domain severely inhibits HIV-1 integration (260). It has been suggested that chromatin-associated LEDGF/p75 acts as a chromosomal receptor for the HIV-1 PIC, tethering it to the chromosome and thereby promoting the strand transfer reaction.

The pattern of chromosomal integration by HIV-1 and other retroviruses was for many years believed to be virtually random. More recently, the occurrence of severe adverse events related to retroviral integration during gene therapy, combined with the availability of methods to clone, sequence, and map integration sites, has led to a greater understanding of retroviral integration site preferences. Retroviral integration is not thought to be strongly dependent upon primary sequence, but weak consensus sequences have been detected at HIV-1 integration sites (261-263). Interestingly, the consensus sequences are palindromic, and the centre of rotation is the centre of the host sequence that is duplicated during integration. The weakness of the consensus sequences suggests that the contribution of primary sequence to HIV-1 integration site

selection is not one of specific recognition of base pairs by integrase, but rather that particular sequences result in local physical properties of DNA such as bendability and protein deformability that are conducive to the mechanism of integration (264).

The position of DNA binding proteins can affect local target site selection. Sequence-specific DNA binding proteins can block integration into their binding site, possibly through steric interference (265). The outward-facing major groove of DNA wrapped around nucleosomes is favoured as a target for integration *in vitro* and *in vivo* (266;267), perhaps because the distortion of the target DNA makes the integration reaction more energetically favourable.

On a larger scale, HIV-1 displays a preference for integration into actively transcribing genes (268). The mechanism by this preference comes about is not well understood (reviewed in (269)). Knockdown of LEDGF/p75 in target cells biases HIV-1 integration away from active genes and towards CpG islands and transcription start sites, suggesting a role for this factor in HIV-1 integration site selection (270). Although there is a high degree of sequence homology between retroviral integrases, gammaretrovirus (271) and foamy virus (272) integrases do not interact with LEDGF/p75 and integrate preferentially near CpG islands and transcription start sites.

Following integration, the presence of a provirus may disrupt nearby host genes or regulatory elements with potentially deleterious effects on host cell function, as described in Section 2.2.

2.3.1.2. Lentiviral Vector Development

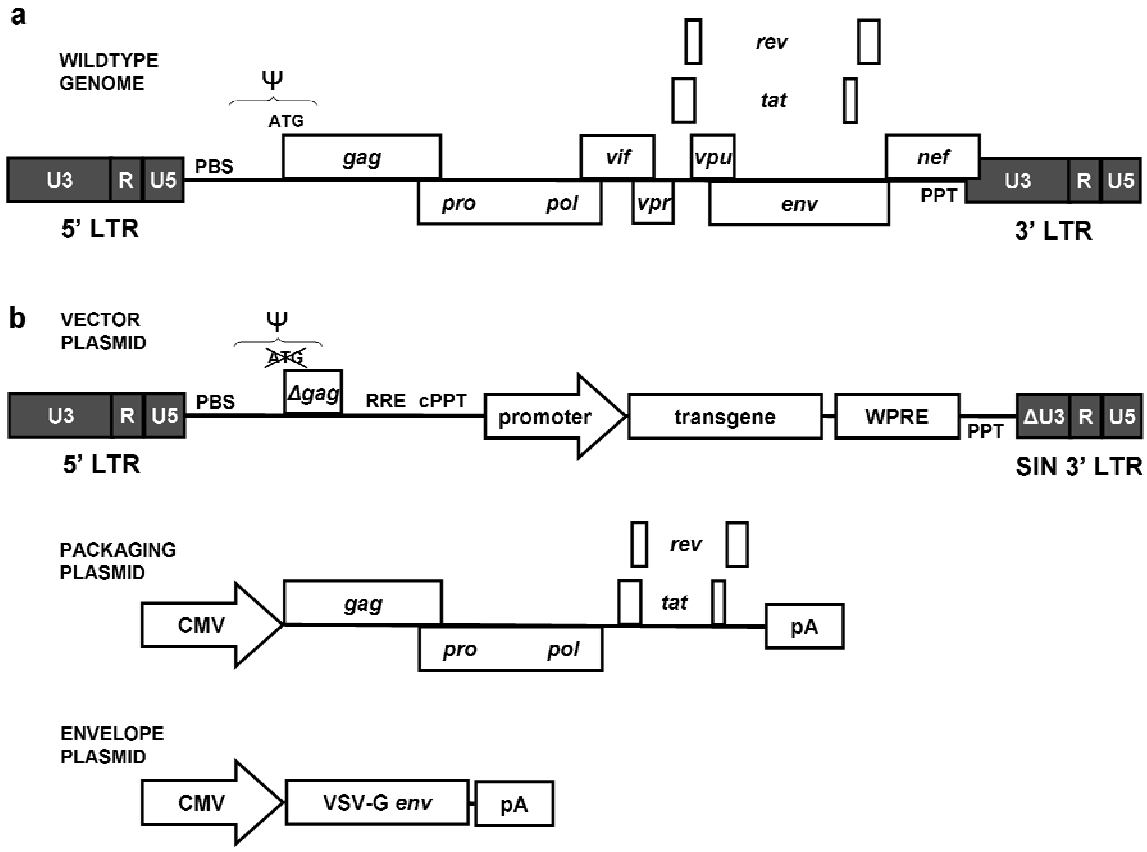

Figure 2.8 HIV-1 genome and second generation lentiviral vectors.

(a) Map of a wildtype HIV-1 provirus. LTR, long terminal repeat (subdivided into U3, R, and U5 regions); PBS, tRNA primer binding site; Ψ, RNA packaging signal; gag, polyprotein encoding virion structural components; pro-pol, polyprotein encoding viral enzymes; vif, vpr, vpu, and nef, accessory genes; rev and tat, regulatory genes; PPT, polyprurine tract. **(b)** Map of a three plasmid second generation lentiviral vector system. Δgag, truncated *gag* gene lacking an ATG start codon; RRE, Rev response element; cPPT, central polypurine tract; WPRE, woodchuck hepatitis virus posttranscriptional regulatory element; SIN 3' LTR, self-inactivating long terminal repeat; ΔU3, self-inactivating 3' U3 region lacking promoter activity; CMV, human cytomegalovirus immediate early promoter; pA, polyadenylation signal; VSV-g *env*, envelope protein from the vesicular stomatitis virus.

Retroviral vectors based on gammaretroviruses were among the earliest viral vectors developed for mammalian gene transfer (273). Lentiviral vectors based on HIV-1 and foamy virus vectors were developed later using many of the principles of the original gammaretroviral systems (274;275).

The largest structural change in moving from virus to vector was to split the genome into the non-coding sequences required *in cis* for gene transfer and the viral coding

sequences required only *in trans* in the producer cell. This separation renders the vector capable of one round of infection only, since no viral proteins will be produced in the target cell. A common split-genome system results in three plasmids: a transfer vector containing the essential *cis* elements and the transgene; a packaging plasmid expressing the essential polyproteins Gag and Gag-Pol and sometimes Tat, Rev, and the accessory proteins for HIV-1 vectors; and a plasmid for expression of the viral envelope. Separation of the genome onto multiple plasmids reduces the risk of recombination resulting in the production of replication-competent retroviruses (RCRs). This was historically a significant safety concern for early retroviral vector production systems which routinely gave rise to replication competent viruses through recombination (276) but has not been observed with modern split-genome systems.

The transfer vector consists of the viral LTRs, the RNA packaging signal (Ψ), and the transgene expression cassette. The viral LTRs contain sequences essential for transcription, reverse transcription, and integration of the vector. The packaging signal extends into the beginning of the *gag* coding sequence, but the start codon of *gag* has been inactivated by mutation. Later improvements to the transfer vector include the insertion of a cPPT element into lentiviral vectors to enhance transduction of nondividing cells (277) and a woodchuck posttranscriptional regulatory element (WPRE) which enhances viral titre and possibly transgene expression (278). It was reported that a short 'X-protein' encoded within the WPRE might be oncogenic (279), but inactivation of X-protein expression does not affect WPRE function (280). Self-inactivating transfer vectors were developed for improved safety in which almost all of the 3' U3 element is removed (281;282) in order to eliminate its promoter/enhancer activity. The mechanism of reverse transcription is such that both proviral U3s originate from the 3' LTR of the RNA genome, so proviruses resulting from infection with this vector lack LTR-driven transcription and cannot transcribe their full genome. Though strong constitutive promoters are commonly used to drive transgene expression, regulated expression from transfer vectors can be achieved to some extent through the use of tissue-specific or inducible promoters or microRNA target sequences (283;284).

The first generation HIV-1 vector packaging plasmid supplied all viral coding sequences except *env* (285). In the second generation, the accessory proteins were

eliminated as it was shown that they were not essential for vector function (286). In the third generation, Tat was eliminated from the packaging system by replacing the 5' U3 promoter of the transfer vector with viral promoters which do not require Tat. Rev is supplied on a fourth plasmid in this system (287). It has been demonstrated that Rev can be removed from the packaging system by stabilising viral transcripts through codon optimisation, but this system has not been widely adopted (288). Non-integrating lentiviral vectors have been developed which contain class I mutations to the integrase catalytic site. These display an integration efficiency which is estimated to be 10^{-4} times that of standard vectors (289).

Retroviral vectors are able to incorporate envelope proteins from a wide range of enveloped viruses if they are co-expressed in producer cells, a phenomenon known as pseudotyping (290). The choice of envelope plasmid used to pseudotype the vector alters the target cell specificity and physical properties of the virion (291). The most commonly used envelope protein is the vesicular stomatitis virus glycoprotein (VSV-G) which confers stability and broad tropism upon virions (292), though vector preparations produced with this envelope may contain immunostimulatory contaminants (293). Interestingly, the receptor for VSV-G is not definitively known (294;295).

The level of expression from integrated retroviral vectors is subject to positional effects whereby adjacent chromosomal elements modulate the level of transgene expression (296). Interestingly, it has recently been reported that integrating lentiviral vectors express more strongly overall than non-integrating vectors (297;298). Positional effects can be reduced through the inclusion of chromatin insulators in the vector backbone (299). Transgene silencing occurs through mechanisms that are not well understood but may include methylation and histone deacetylation as part of host defences against transposable element expansion (300).

Vector systems based on other lentiviruses have also been reported, including human immunodeficiency virus type 2 (HIV-2), simian immunodeficiency virus, feline immunodeficiency virus, equine infectious anaemia virus, and others (reviewed in (301-303)).

2.3.2. Adeno-Associated Virus Vectors

The adeno-associated virus (AAV) is a single-stranded, non-enveloped DNA virus belonging to the family *Parvoviridae*. Unlike retroviruses, AAV does not require chromosomal integration for progression through the cell cycle. Perhaps as a consequence, the efficiency of AAV integration is considerably lower than that of retroviruses. Nonetheless, two types of AAV integration are known. In the presence of AAV large replication protein Rep78/68 (Rep), wildtype AAV integrates preferentially into a site on human chromosome 19 known as AAVS1 (304). In recombinant AAV vectors lacking Rep, integration occurs more randomly throughout the genome.

The large Rep proteins are essential non-structural proteins with helicase and strand- and sequence-specific endonuclease activities (305). During replication of the AAV DNA genome, Rep binds to a $(GAGC)_3$ Rep binding element (RBE) present in the virus terminal repeats (TR) and nicks the double-stranded DNA genome at a terminal resolution site (trs). An RBE is also present at the AAVS1 site on chromosome 19, and Rep is able to tether the AAV genome to the AAVS1 site (306). However, most intracellular wildtype AAV genomes do not integrate. The mechanism of site-specific integration is not known, but is likely to be complex given that it frequently results in multiple insertions and significant and unpredictable rearrangements to both vector and host DNA (reviewed in (307)).

Commonly used recombinant AAV (rAAV) vectors for gene therapy do not carry the Rep protein and persist in target cells largely as episomal DNA, initially as small circles but gradually developing into large concatemers over time (308). As with wildtype AAV, rAAV integration takes place through a complex and poorly understood mechanism which results in significant rearrangement of host and vector DNA (309). Large scale rAAV integration site analysis revealed integrations throughout the genome, with a weak preference for integration near CpG islands or within transcription units. Interestingly, 8% of integration sites were found within ribosomal DNA (310).

Although initially considered to be an integrating vector, the current view of rAAV vectors is that they are principally non-integrating but may integrate at low efficiency.

It has been estimated that the efficiency of both wildtype AAV site-specific integration by Rep and rAAV non-targeted integration is approximately 10^{-3} integrations per intracellular vector genome (311), two orders of magnitude less than that of lentiviral vectors.

2.3.3. Sleeping Beauty: A DNA Transposon Vector

Sleeping Beauty is a Tc1/*mariner*-type DNA transposon that was reconstructed by consensus alignment of inactive transposon "fossils" from a number of salmonid fish (312) and has been explored as an integrating vector for transgenesis in many eukaryotic species. The complete wildtype transposon consists of a transposase protein coding sequence flanked by two nonidentical 230bp inverted repeats (IRs). A 32-34bp transposase binding site is present at the inner and outer end of each IR (313). The 5' untranslated region between the left IR and the transposase coding sequence is able to promote weak transcription of transposase. The right IR also has weak promoter activity directed towards the centre of the transposon, raising the possibility of double-stranded RNA formation and transposon silencing by RNAi (314;315).

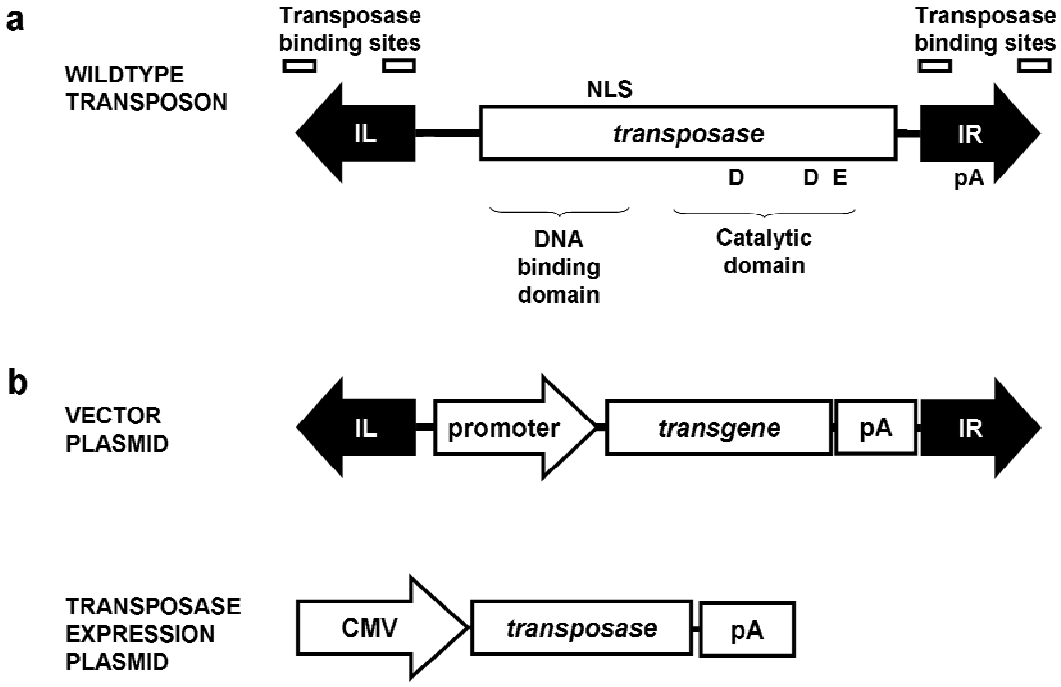

Figure 2.9 Genetic maps of the Sleeping Beauty transposon and vector system.

(a) Wildtype Sleeping Beauty transposon consisting of flanking inverted repeats and a transposase coding sequence. Adapted from (316). **(b)** Two plasmid Sleeping Beauty vector consisting of a vector plasmid containing a transgene expression cassette flanked by inverted repeats and a transposase expression plasmid. IL, left inverted repeat; IR, right inverted repeat; NLS, nuclear localisation signal; DDE, integrase catalytic motif; pA, polyadenylation signal; CMV, human cytomegalovirus immediate early promoter.

The 340 amino acid transposase protein consists of an N-terminal, pairlike DNA binding domain with two helix-turn-helix motifs, an overlapping nuclear localisation signal, and a C-terminal DDE-type catalytic domain which is common to integrases and transposases from many mobile elements, including retroviruses (317;318). Transposase catalyses the excision of the transposon from flanking DNA and its reintegration elsewhere in a 'cut-and-paste' process known as transposition.

Transposition initiates with binding of transposase protein molecules to the two binding sites present within each IR (i.e. four sites in total) (319). Binding takes place more strongly at the inner than outer sites (320). It has been reported that transposase displays a high binding affinity for transposon DNA which is packaged into heterochromatin (321;322). It has also been reported that linearised, non-supercoiled DNA is 15-20 fold less efficient as a substrate for transposition (323;324). After binding, the four transposase proteins form a tetramer, bringing the two IRs into close proximity in a structure known as the synaptic complex (325). This process requires the presence of the host DNA-bending protein HMGB1 (326) and is enhanced by an 11bp 'half-direct repeat' transpositional enhancer present only in the left IR (327).

Once the synaptic complex has formed, the transposase catalytic domain cleaves the DNA at the IR ends to excise the transposon. Cleavage is dependent upon the presence of flanking TA dinucleotides and is enhanced when the transposon is flanked by TATA motifs (328;329). Cleavage is staggered so that the excised transposon possesses 3 nucleotide 3' overhangs at both ends (330). The excised transposon subsequently integrates into a new target DNA at a TA-dinucleotide. DNA repair pathways such as non-homologous end joining (NHEJ) repair the donor and target DNA strands, resulting in a 3bp transposon footprint on the donor DNA and duplication of the TA-dinucleotide on the target DNA (331;332). Sleeping Beauty transposase interacts with the NHEJ repair pathway component Ku, and Ku is necessary for efficiency and fidelity of the transposition reaction (333;334).

Transposase also interacts with Miz-1, and this leads to downregulation of cyclin D1 expression and subsequent extension of the G1 phase of the cell cycle (335). NHEJ activity is high during G1, so it is likely that this manipulation of the cell cycle is a selfish transposon strategy to enhance its transposition. This may be a general strategy for mobile parasitic elements, as HIV-1 also manipulates the cell cycle to its advantage through the action of Vpr (336;337).

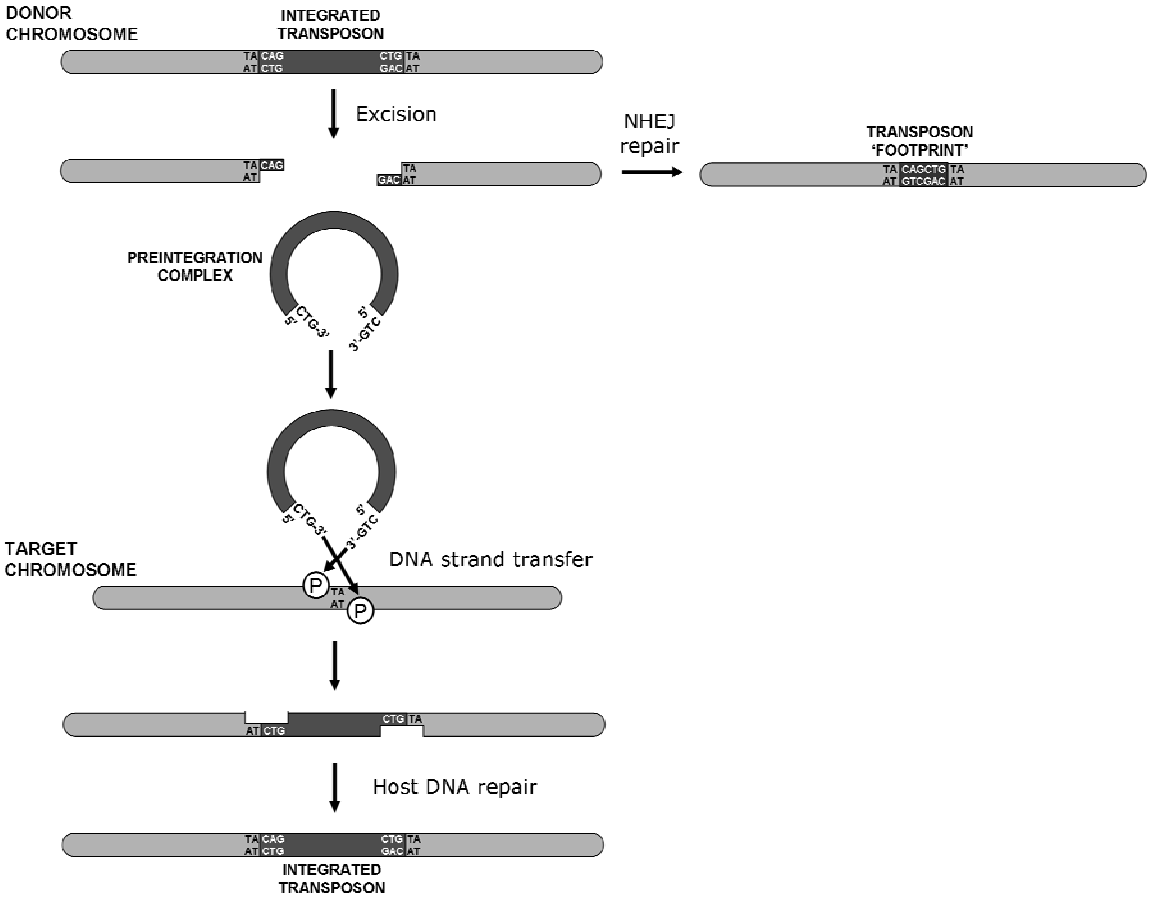

Figure 2.10 The mechanism of Sleeping Beauty transposition.

An integrated transposon is excised by molecules of the transposase protein, leaving the donor chromosome with 3bp transposon-derived overhangs. These are repaired by non-homologous end-joining (NHEJ), resulting in a transposon 'footprint'. The excised transposon remains bound to transposase as a preintegration complex. When this encounters a suitable target chromosome containing a TA dinucleotide, a strand transfer reaction inserts the transposon at this position. Filling of single stranded DNA by host DNA repair proteins results in a duplication of the target TA dinucleotide at each end of the transposon integration site. P, phosphate group in DNA backbone. Adapted from (338).

As with retroviruses, Sleeping Beauty transposons are able to integrate throughout the genome. Sleeping Beauty reintegration occurs preferentially on the same donor molecule (including plasmid donors), a phenomenon known as 'local hopping' (339). Potential target TA-dinucleotides differ in their attractiveness for integration due to local DNA properties such as DNA deformability, and algorithms have been constructed which predict this attractiveness (340). A weak consensus sequence at integration sites was found to be 5'-RCAYA**TA**TRTGY-3' (341). Integration site analysis in cell lines has shown that integration of Sleeping Beauty is almost random with respect to genomic features such as transcription units and CpG islands (342;343).

The ability of Sleeping Beauty to stably integrate into host chromosomes has led to its development as a vector, principally as a means of integrating plasmid DNA delivered with non-viral vectors. In the first generation vector, pT, the transposon was split into two plasmids; the transposon vector plasmid contains a transgene expression cassette flanked by IRs, while the transposase plasmid drives expression of transposase. When both plasmids are introduced into target cells, transposase expression *in trans* enables transposition of the transposon from the vector plasmid into a host chromosome (344). In most systems transposase expression has been provided on a separate plasmid, but single vector systems carrying both transposon and transposase (345) and direct delivery of transposase mRNA (346) have also been reported. The level of transposase expression in target cells must be tightly controlled, as overexpression of transposase actually reduces the level of transposition in a phenomenon known as overproduction inhibition (347). The mechanism of overproduction inhibition is not known, but it has been reported in other Tc1/*mariner* transposons and may represent an autoregulatory system (348).

The efficiency of transposition from the transposon vector plasmid declines with transgene size at an exponential rate of about 30% per additional kb, limiting its useful capacity (349;350). The efficiency of transposition was improved in the second generation vector pT2, which used IRs with a better fit to the known Tc1 consensus sequence to give a 4-fold enhancement of transposition (351). Although it was not realised at the time, this vector reconstruction also removed transposon sequences immediately flanking the IRs which are now known to promote weak transcription

into the transposon centre (352;353). A third generation vector plasmid, pT3, was constructed in which the right IR was replaced with an inverted copy of the left IR, resulting in a 2-fold increase in transposition (354). However, unlike pT2, pT3 contains sequences adjacent to the IRs now known to have weak promoter activity.

The Sleeping Beauty transposase has also undergone several generations of development through fitting to transposase consensus sequences and alanine scanning mutagenesis to identify residues which alter transposition activity (355-359). The first transposase, SB10, was able to confer resistance to the antibiotic G418 in 0.7% of HeLa cells following lipid-based co-transfection with the pT vector, while an improved transposase, SB11, resulted in 2% resistance (360). Codon-optimisation of SB10 to improve expression in human cells did not increase transposition activity (361). HSB17, an enhanced, codon-optimised transposase, produced 8.5% G418-resistant HeLa cells following lipid-based transfection with the pT vector (an SB10 control yielded 0.45% resistance in this experiment) (362). A recently developed hyperactive transposase, SB100, has been reported to be 100-fold more active in transposition than SB10 while retaining a virtually random integration profile (363).

It has recently been suggested that Sleeping Beauty insertions are subject to a high degree of gene silencing. Garrison *et al* reported that as few as 1 in 15 chromosomal insertions in cell lines were able to express a fluorescent reporter gene, though treatment of cells with methyltransferase or histone deacetylase inhibitors raised this proportion to almost 100% (364). Similarly, Dalsgaard *et al* reported a 15-fold increase in the number of cells expressing an antibiotic reporter when chromosomal insulator elements were included within the transposon, though the rate of transposition was unaffected (365). The mechanism of silencing is not known, though a significant proportion of insertions are methylated (366). Tc1 transposon silencing can take place through RNAi following the production of double stranded RNA specific to the transposon IRs (367).

Sleeping Beauty vectors have been tested in pre-clinical gene therapy models in tissues including liver (368;369), lung (370), and tumour (371). A clinical trial proposal has been presented to the NIH Recombinant DNA Advisory Committee which plans to use electroporated Sleeping Beauty plasmids to express an anti-CD19 chimeric antigen receptor in autologous T-cells for immunotherapy of $CD19^+$ B-

lymphoid malignancies (372;373). However, approval of the trial has been postponed pending further studies into the risk of Sleeping Beauty integration causing insertional mutagenesis and unregulated T-cell expansion. To date, no case of oncogenic insertional mutagenesis has been reported in normal or cancer-prone animals treated with standard Sleeping Beauty vectors (374).

Following the reconstruction of Sleeping Beauty, a number of other DNA transposons were investigated as integrating vectors in mammalian cells, including a second Tc1-like transposon from the Northern leopard frog *Rana pipens* designated Frog Prince (375), the piggybac transposon from the cabbage looper moth *Trichoplusia ni* (376), and the hAT-like transposon Tol2 from the Japanese medaka fish *Oryzias latipes* (377). The piggybac transposon is reported to have a larger carrying capacity than Sleeping Beauty, up to 14kb without loss of transposition efficiency (378). Piggybac does not leave transposon footprints following excision and is not subject to overproduction inhibition, but is reported to integrate into transcription units more frequently than Sleeping Beauty (379). In direct comparisons in cell lines, piggybac was found to transpose approximately 2-fold more efficiently than an SB11-pT2 system, resulting in 0.9% of HeLa cells becoming hygromycin resistant compared to 0.4% with Sleeping Beauty.

2.3.4. Tyrosine and Serine Recombinases

Recombinases are specialised proteins which catalyse site-specific recombination between short recognition sites present on two DNA molecules. Many recombinases have been identified in a large number of prokaryotes and eukaryotes, but almost all fall into two families, namely the tyrosine and serine recombinases.

The tyrosine recombinase family (also known as λ integrases) include the Cre recombinase from bacteriophage (380) and Flp recombinase from the yeast *Saccharomyces cerevisiae* (381), while the serine recombinase family includes phiC31 integrase from bacteriophage (382;383). The mechanism of recombination used by each family has been well-studied (reviewed in (384;385)). Although the two are mechanistically quite distinct (for example, tyrosine recombinases establish a Holliday junction during recombination while serine recombinases do not), there are similarities between the two recombination processes.

Both tyrosine and serine recombination sites contain two inverted recombinase binding sites. The total recombination site size varies between recombinases but is greater than 30bp for the most commonly used systems. During recombination, the recombination site on each DNA molecule is bound by recombinase enzymes. Serine recombinases introduce one nick into each of the four DNA strands, while tyrosine recombinases nick only one strand on each parental DNA. Resolution of the nicked intermediates results in recombination of the two parental DNA molecules. When at least one parental molecule is circular (e.g. a plasmid) and the other is linear (e.g. a chromosome), the net result of recombination is to insert the circular parent into the recombination site of the linear parent.

A number of recombinases have been investigated for use in gene therapy, principally the Cre and Flp recombinases and phiC31 integrase (386). A significant limitation is the requirement for target recombination sites in genome. Although human cells contain no perfectly matched recombination sites for any of these three enzymes, divergent pseudo-attP sites are present that can be used by phiC31 integrase with reduced efficiency (387). A survey of the integration sites used by phiC31 integrase in human cell lines found that several hundred potential sites exist, but the majority of integrations take place at a small subset of these (388). There are no reports of oncogenic transformation due to phiC31 integrase expression *in vivo*, but safety concerns have been raised following reports that the enzyme can cause chromosomal rearrangements in mammalian cells (389;390).

In terms of efficiency, transfection experiments using phiC31 integrase expression and donor plasmids resulted in 0.1% stably G418-resistant HeLa cells (391). An enhanced phiC31 integrase has recently been developed which is 2-fold more efficient (392).

2.3.5. Site-Specific Nucleases for Gene Targeting

In gene targeting, a DNA fragment introduced into cells is able to replace a portion of endogenous chromosomal DNA through homologous recombination (HR). For example, mouse embryonic stem (ES) cells can be transfected with partially homologous template DNA in order to produce specific genomic alterations (393). This has contributed greatly to basic biological research as it enables the production

of adult mice carrying specific genomic alterations (394). Gene targeting is a highly attractive approach to gene therapy as it offers the potential to insert therapeutic DNA at a known, safe location, or even to correct disease-causing mutations *in situ*. However, gene targeting in mammalian cells is extremely inefficient, with just 1 in 10^6 mouse ES cells carrying the desired insertion and 100 to 1000-fold more carrying background integrations elsewhere in the genome.

It was shown first in yeast (395) and later in mammalian cells (396) that the efficiency of gene targeting could be enhanced by the introduction of a double strand DNA break (DSB) at the target site using a site-specific endonuclease. The free ends at DSBs mark them as substrates for host DNA repair pathways, stimulating HR between the target and template DNA (reviewed in (397)). Rouet *et al* reported that the expression of a targeting endonuclease increased the rate of gene targeting 100-fold (398).

DSB repair can produce several outcomes. Repair via non-homologous end joining (NHEJ) may restore the original sequence, resulting in no change at the target site. NHEJ may also introduce small insertions of several base pairs. If a template DNA containing homology to the sequence flanking the DSB is available, homologous recombination (HR) repair may result in the replacement of DNA surrounding the DSB with DNA from the template. Competition between the NHEJ and HR pathways for DNA repair is a recognised problem for gene targeting approaches.

A number of endonucleases have been employed for mammalian genome editing. Hundreds of naturally-occurring meganucleases have been identified which have recognition sites at least 12bp long, a length usually sufficient to be unique in mammalian cells. However, as with recombinases, none of these enzymes have useful target sites in human cells, and retargeting their substrate specificity by protein engineering has proven extremely difficult (reviewed in (399)).

The ability to engineer proteins which bind specifically to useful DNA target sites has long been a goal in biotechnology. A possible solution was offered by the elucidation of the structure of the DNA binding domain from a zinc finger transcription factor (400). This showed a relatively simple recognition motif in which a series of looped polypeptides, or ‘fingers’, contact three DNA bases per finger using three amino acids

within each finger. The structure suggested that zinc finger transcription factors were in fact modular, and that the fingers could be interchanged to alter the DNA binding site specificity at will. The 'modular assembly' approach stimulated the collection of libraries of fingers, each finger recognising a different DNA triplet (401;402). Construction of artificial zinc finger domains proceeds by assembly of fingers to give the desired binding specificity followed by *in vivo* selection of functional domains (403). It has been shown that endonuclease domains such as that of FokI can be fused to zinc finger DNA binding domains to give zinc finger nucleases (ZFNs), artificial endonucleases which can be targeted against an extremely large number of chromosomal sites (404;405). ZFNs are not thought to be more efficient than natural endonucleases at stimulating gene targeting, but their flexibility of specificity has raised the prospect of useful applications.

Using ZFNs and template DNA delivered by plasmid transfection, Urnov *et al* were able to introduce small sequence changes into the endogenous IL2RG locus in 20% of K562 cells and 5% of primary human $CD4^+$ T lympocytes (406). It was later shown that up to 8kb of heterologous sequence can be inserted into this locus in cell lines by flanking it with 750bp arms homologous to the sequence surrounding the DSB (407). Gene addition of this type was shown at the CCR5 locus in 39% of Jurkat cells and 0.06% of primary $CD34^+$ haematopoietic progenitor cells (408).

The insertion of base pairs during NHEJ can introduce frame-shift mutations if the DSB occurs within an exon. This strategy has been exploited to knock out expression of CCR5, a major co-receptor for HIV-1 infection of T lymphocytes, in order to protect these cells from HIV infection (409). In February 2009, Sangamo Biosciences commenced a clinical trial of this system.

One concern with the application of nucleases for gene addition or knockout is the cytotoxicity observed with intracellular endonuclease expression (410;411). Overexpression of the meganuclease I-SceI from an adenovirus 5 vector was fatal in 40% of mice (412). In general, endonuclease toxicity is thought to be due to off-target DNA cleavage and is generally greater for ZFNs than meganucleases. This issue is being addressed by the development of ZFNs with improved target specificity (413;414).

2.3.6. Group II Introns

Mobile group II introns are transposable elements found in a number of prokaryotes and the mitochondria and chloroplasts of some eukaryotes (reviewed in (415)). The integrated DNA-form is transcribed into RNA which is able to integrate elsewhere on prokaryotic chromosomes. Uniquely, integration involves *in situ* reverse transcription at the integration site, a process which is catalysed by an intron-encoded protein (IEP) with reverse transcriptase activity and significant homology to retroviral reverse transcriptases (416). The target site specificity of group II introns is determined through base pairing between 14 nucleotides of the intron RNA with the target site DNA; thus, integration site selection can be reprogrammed simply by editing these nucleotides, subject to few constraints. In addition, it is possible to insert heterologous sequences such as drug resistance markers into the intron (417). This high degree of flexibility has led to the development of the L1.LtrB group II intron from *Lactococcus lactis* as a programmable 'targetron' vector for site-specific integration in prokaryotes (418). Expression of a targetron vector in *E. coli* results in chromosomal integration in approximately 1% of cells and target site specificity is greater than 98% (419).

L1.LtrB has also been investigated for activity in eukaryotic cells. Guo *et al* generated group II introns targeted against human CCR5 or HIV-1 pol and confirmed integration into target plasmids carrying these sequences in *E. coli*. Group II introns were then reconstituted *in vitro* using purified IEP and RNA and transfected into human cell lines together with the target plasmids. Site-specific integration into the target plasmids was detected by PCR (420). Recently, Mastroianni *et al* injected reconstituted introns into embryos of the frog *Xenopus laevis* or fly *Drosophila melanogaster* and demonstrated integration into target plasmids in approximately 3% of these cells in the presence of increased Mg^{2+}. Integration into eukaryotic chromosomes by a group II intron was demonstrated for the first time at the *yellow* gene in *Drosophila*, with successful targeting confirmed by PCR and sequencing of the intron-chromosome junction (421). Although the efficiency of the reaction was low, this study demonstrated the principle that targeted integration into eukaryotic chromosomes can be achieved without the need for protein engineering.

2.3.7. Fusion Proteins to Retarget Integration

The risk of insertional mutagenesis following semi-random integration of retroviruses or transposons makes retargeting integration to a particular chromosomal location an attractive goal. Retargeted integration would insert vector DNA in a predictable location and so reduce positional effects on transgene expression, as well as reducing the risk of insertional mutagenesis due to proto-oncogene or tumour suppressor gene disruption. The most common approach to retargeting an existing semi-random integration mechanism is to fuse elements of the preintegration complex to sequence-specific DNA-binding domains (DBDs), thus tethering the complex to a desired chromosomal location where the integration mechanism can proceed.

Two approaches have been taken to retargeting lentiviral vector integration. Ciuffi *et al* engineered an artificial tethering factor consisting of the integrase binding domain of LEDGF/p75 fused to the DNA-binding domain of the phage λ repressor. An *in vitro* integration reaction was performed using purified tethering factor and HIV-1 integrase to catalyse integration into DNA containing the λ repressor target site. The presence of the tethering factor was found to increase the rate of integration at sites surrounding the target site (422). This effect has not yet been reported in cells. A second approach has been to fuse the HIV-1 integrase protein directly to a site-specific DBD. This strategy was tested in cell lines using lentiviral vectors carrying an E2C-IN fusion protein targeted against the erbB2 site on chromosome 17. 1.5% of integration sites were found to lie near the erB2 target site in cells transduced with E2C-IN vector compared to 0.15% of sites in cells transduced with standard IN vector, an enrichment of almost 10-fold (423).

Efforts have been made to target Sleeping Beauty integration to specific chromosomal target sites by constructing fusion proteins made up of the transposase and a sequence-specific DNA binding domain. C-terminal additions to the transposase appear to completely abolish activity, and even small N-terminal additions of 6 amino acids significantly reduce transposition efficiency (424;425). A transposase-DBD fusion protein targeted by Yant *et al* to the E2C site on chromosome 17 was able to enrich the frequency of integration close to a plasmid target site 10-fold compared to the expected value with an unmodified transposase. No increase in integration close to

the chromosomal target was observed (426). A similar study from Ivics *et al* reported no success with transposase-DBD fusions but was able to show targeted integration into chromosomes using a tethering factor consisting of a DBD specific for an artificial chromosomal tetO target and a protein-binding domain specific for the transposase protein (427). This factor was able to target 10-25% of all transposition events into TA-dinucleotides within 2kb of the tetO binding site, an enrichment of 10^7-fold relative to standard transposition.

Retargeting an existing integration mechanism is attractive as it offers the potential to retain the high efficiency of integration observed with the native pathway. However, the approaches attempted to date have been hampered by the fact that the native pathway remains able to integrate independently of the added retargeting system. The fact that off-target DNA is present in great excess relative to target DNA means that integrations at the target site usually constitute only a small proportion of the total integration events.

2.3.8. Hybrid Vectors

An alternative approach to retargeting integration is to provide a given vector with an entirely new integration mechanism. A number of such hybrid vectors have been investigated which combine the gene transfer activity of one vector with the integration activity of another (Table 2.1). This combination is particularly useful where either the gene transfer vector is normally non-integrating or the integration system is unable to move between cells. In this way it is possible to generate vectors which combine desirable cell tropism and integration properties for particular gene therapy applications.

		Gene transfer vector				
		Lentivirus	Herpes virus	Adenovirus	AAV	Non-viral
Integration method	Sleeping Beauty transposase	This study (428) Staunstrup 2009 (429)	Bowers 2006 (430)	Yant 2002 (431)		Ivics 1997 (432)
	Retro/lentiviral integrase	Naldini 1996 (433)		Murphy 2002 (434)		Shoji-tanaka 1994 (435)
	AAV Rep	Philpott 2007 (*pers. comm.*)	Heister 2002 (436) Wang 2002 (437)	Recchia 1999 (438)	Zhang 2007 (439)	Surosky 1997 (440)
	Cre/Flp recombinase	Moldt 2008 (441)				Sauer 1990 (442) Koch 2000 (443)
	phiC31 integrase	Thornhill 2005 (*pers. comm.*)		Ehrhardt 2007 (444)		Groth 2000 (445)
	Meganuclease	Cornu 2007 (446)			Miller 2003 (447) Porteus 2003 (448)	Rouet 1994 (449)
	Zinc finger nuclease	Lombardo 2007 (450)				Urnov 2005 (451)

Table 2.2 Hybrid vector publications.

Literature search performed April 2009.

Hybrid vectors require the nucleic acid output from the gene transfer step to be a compatible substrate for the integration step. In this study we have generated a hybrid lentivirus-Sleeping Beauty transposon vector, so the remainder of this section will review the literature regarding hybrid vectors containing either a lentiviral vector or Sleeping Beauty component.

Three hybrid lentiviral vectors have been published in which the endogenous HIV-1 integrase activity was eliminated through the introduction of the Class I mutation D64V and integration was instead carried out by another mechanism.

Firstly, Cornu *et al* reported a lentiviral vector able to undergo gene targeting by homologous recombination stimulated by the meganuclease SceI. In this study, an SceI expression cassette and the homologous repair template were cloned into two separate non-integrating lentiviral vectors (NILVs) and co-transduced into target cells. Gene conversion at a chromosomal eGFP target was observed in approximately 1% of HEK293 cells transduced with both NILVs and 0.03% of cells transduced with the template NILV only. Interestingly, gene conversion occurred in 12% of HEK293 cells when SceI was expressed from an integrating lentiviral vector (ILV). The authors suggest that low expression from NILVs (also reported elsewhere e.g. (452)) was a limiting factor in the gene targeting frequency (453).

Secondly, Lombardo *et al* also generated NILVs able to integrate by homologous recombination. In this study, homologous recombination was stimulated by zinc finger nucleases (ZFNs). Target cells were transduced with three NILVs, two for expression of each half of the ZFN dimer and one for delivery of the homologous repair template. Gene conversion was observed at the IL2RG locus in 16% of K562 cells transduced with all three vectors. The efficiency of conversion in the absence of the ZFN was not reported, but appears to have been low. As well as gene conversion, gene addition into the IL2RG locus in K562 cells was achieved with an efficiency of 3.4% by flanking an eGFP expression cassette with arms homologous to the IL2RG. In the absence of the ZFN, 1.1% of cells became stably eGFP positive. Gene addition at the CCR5 locus occurred in 35% of K562 cells with a 2% background integration rate and 0.06% of $CD34^+$ haematopoietic progenitor cells with a 0.005% background integration rate (454). The authors argue that the low rate of gene targeting in

haematopoietic progenitor cells is due to the low efficiency of transduction commonly observed with these cells (455).

Thirdly, Moldt *et al* presented a proof-of-principle vector in which cells were co-transduced with two NILVs, one for expression of an enhanced Flp recombinase and the other providing a hygromycin resistance cassette adjacent to an FRT (Flp recombination) site. The mechanism of Flp-directed recombination is such that only circularised NILV molecules were able to insert as desired, but it is known that approximately one third of linear lentiviral DNA forms 1- and 2-LTR circles in target cells through the action of host DNA repair proteins (456). Co-transduction of an HEK293 cell line engineered to contain 3 chromosomal FRT sites resulted in stable expression of hygromycin resistance in 0.01% of cells compared to 0.4% of cells co-transfected with plasmids carrying the same Flp expression and FRT-hygromycin resistance cassettes. No background integration was observed because the vector was designed so that hygromycin resistance could only be expressed following FLP-mediated insertion at the target site. Sequencing of insertions confirmed that 1- and 2-LTR circles were the substrate for Flp-mediated insertion. The authors used quantitative PCR to determine the efficiency of transfection or transduction, detecting copy numbers of 1.3 x 10^4 plasmids per cell and 22 2-LTR circles per cell 24 hours after gene transfer. They thus estimated the rate of integration due to Flp to be 2.8 x 10^{-7} insertions per plasmid versus 1.7 x 10^{-5} per 2-LTR circle. The authors speculate that the greater efficiency of 2-LTR insertion per intracellular copy could be due to differences in subcellular localisation (e.g. improved nuclear entry of NILVs) or structural differences in the substrate DNA such as supercoiling.

Three hybrid vectors based on Sleeping Beauty have been published. The first was the use of non-viral transfection to deliver Sleeping Beauty plasmids to target cells (457). Non-viral transfection of plasmid DNA is the gold standard for Sleeping Beauty delivery, and relatively little research has addressed the suitability of transfected plasmid DNA to act as a substrate for transposition. Yant *et al* reported that transposition from linearised plasmid DNA is 20-fold less efficient than transposition from circular plasmids (458). It is important to note that this study did not control for differences in the efficiency of transfection between linear and circularised plasmid DNA. While supercoiling is known to influence the efficiency of transposition in a

number of mobile elements, its importance for the Tc1/*mariner* transposon family is controversial and the requirement does not seem to be absolute (459;460). It was recently reported that experimentally-induced heterochromatinisation of Sleeping Beauty transposons integrated into chromosomes enhanced excision 100-fold, but this effect has not been demonstrated in plasmid systems (461).

Yant *et al* reported the development of a hybrid Sleeping Beauty-adenoviral vector for *in vivo* gene delivery to mouse liver (462). When an adenoviral two vector system (one each for delivery of the transposon and transposase expression cassette) was delivered to mouse livers by tail vein injection, there was no difference in the long term level of the alpha-1 antitrypsin transgene product in blood plasma. On the basis of the linear versus circular plasmid transfection experiment described above, the authors decided to circularise the transposon vector *in situ* by adding the Flp transgene to the transposase expression vector and flanking the transposon with FRT sites. 45% of mouse hepatocytes were found to contain circularised vector genomes 5 weeks after the initiation of Flp expression compared to 2% in the absence of Flp. When hepatocytes were induced to divide rapidly by administration of CCl_4 to the transduced mice, the level of transgene persistence was estimated to be 11-fold higher in livers transduced with the active transposase. In addition, 7 chromosomal transposon integration sites were successfully recovered from these mice. The two adenoviral vector Flp circularisation system was tested in long term expression experiments in mouse livers, and significantly greater expression from a transposon-borne coagulation factor fIX was observed at 23 weeks in mice transduced with an active transposase, though this level was 0.3% of the initial expression level achieved immediately after transduction. Finally, the transposon, Flp, FRT, and transposase elements were combined on a single adenoviral vector, and efficacy was demonstrated by the recovery of 7 integration sites from mouse liver transduced *in vivo*. The authors did not report integration efficiency data obtained in cell lines.

Lastly, Bowers *et al* generated a hybrid Sleeping Beauty-herpes simplex virus vector in which a lacZ-neo transposon and a transposase expression cassette were carried on two separate vectors. Co-transduction of the BHK cell line demonstrated 10-15% stable G418 resistance in the presence of transposase compared to approximately 0.6% background integration. In primary cortical neuron cultures, co-transduction

with an active transposase resulted in a 2-fold increase in the number of β-galactosidase positive cells (the total cell number was not reported) and a 20-fold increase in vector copy number at days 4 and 9 post-transduction. Finally, mouse embryonic brains were injected with the hybrid vectors 1 week before term and the animals were sacrificed 90 days after birth. Significant β-galactosidase staining was observed in large regions of brains co-transduced with the transposon and an active transposase, and immunofluorescence microscopy revealed expression in a range of cell types. No β-galactosidase staining at all was observed in brains containing the transposon vector but no transposase. 4 chromosomal transposon integration sites were successfully recovered from brains co-transduced with transposase.

The successful development of six hybrid vector systems containing either a non-integrating lentivirus or Sleeping Beauty transposon component suggests that compatibility between the gene transfer vector and the downstream integration mechanism is not a common problem for either of these systems and supports the feasibility of a hybrid lentivirus-transposon vector.

2.4. Summary and Aims

Successful clinical trials using gammaretroviral vectors have highlighted the usefulness of integrating vectors for gene therapy, particularly in mitotic tissues. However, the semi-random nature of retroviral integration introduces a risk of insertional mutagenesis, potentially leading to cell transformation and serious consequences for patients. The development of retroviral vectors with safer patterns of integration site selection would be an attractive solution to this problem.

Retroviral vectors derived from gammaretrovirus, foamy virus, and lentivirus are able to transfer genes efficiently to a broad range of target cell types. However, gammaretroviral and foamy virus vectors integrate preferentially near transcription start sites and CpG islands, while lentiviral vectors integrate preferentially into active genes. In all three cases, the pattern of integration is biased towards important host genomic elements.

By contrast, the Sleeping Beauty DNA transposon has an almost random integration profile with respect to underlying genomic features. It has been suggested that this

pattern of integration may theoretically reduce the risk of insertional mutagenesis. A significant hurdle to the wider application of Sleeping Beauty to gene therapy is the inability of the vector to enter target cells unaided. A common approach has been to transfect Sleeping Beauty plasmids into target cells using non-viral vector systems, but this is not practical for important targets such as haematopoietic stem cells where non-viral transfection is inefficient (463).

Thus, the aim of this study was to develop a hybrid lentivirus-Sleeping Beauty transposon vector which would combine the efficient cell and nuclear entry properties of lentiviral vectors with the theoretically safer integration profile of Sleeping Beauty.

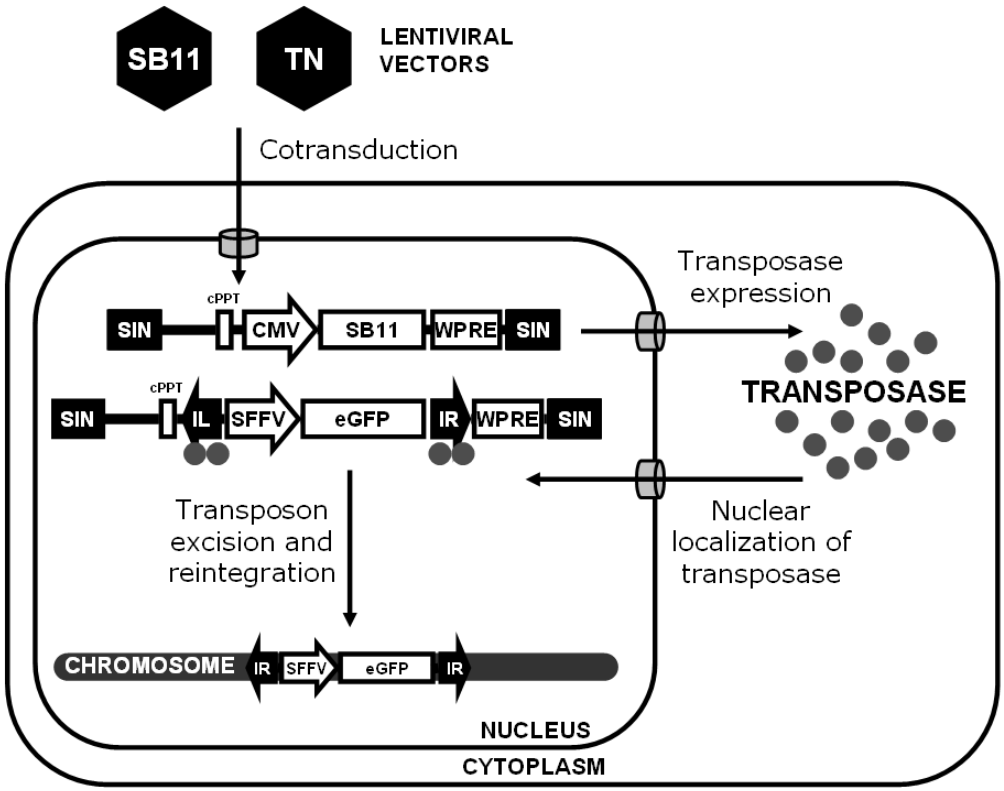

Figure 2.11 Overview of the lentivirus-transposon vector.

TN, Sleeping Beauty transposon; SB11, hyperactive Sleeping Beauty transposase; SIN, self-inactivating (U3-deleted) HIV-1 long terminal repeat; cPPT, central polypurine tract; CMV, immediate early promoter of human cytomegalovirus; WPRE, woodchuck posttranscriptional regulatory element; IL, Sleeping Beauty transposon left inverted repeat; IR, Sleeping Beauty transposon right inverted repeat; SFFV, spleen focus forming virus U3 promoter; eGFP, enhanced green fluorescent protein.

3. Materials and Methods

3.1. Materials

Unless otherwise stated, tissue culture reagents were supplied by Gibco BRL (Invitrogen), general chemicals were supplied by Sigma, and recombinant enzymes for plasmid subcloning were supplied by Promega. DNA sequencing was performed by Functional Biosciences.

3.1.1. General Reagents and Enzymes

1kb Plus DNA ladder	Invitrogen
NlaIII restriction endonuclease	New England Biolabs
Proteinase K	VWR
Agarose	Invitrogen
Agar	MERCK
Ampicillin	Stratagene
G418	Invivogen
TRIzol Reagent	Invitrogen
DNase I, Amplification Grade	Invitrogen
M-MLV Reverse Transcriptase	Invitrogen
Random Primers	Invitrogen

3.1.2. Plasmids

pLNT/SFFV-eGFP-WPRE is a previously described variant of the pHR second generation HIV-1-based vector containing a self-inactivating 3' LTR, a central polypurine tract (cPPT), the spleen focus forming virus U3 promoter (SFFV), an

enhanced green fluorescent protein reporter gene (eGFP), and the woodchuck posttranscriptional regulatory element (WPRE) (464). pLNT/SFFV-MCS-WPRE uses the same backbone but carries a multiple cloning site instead of the eGFP coding sequence.

The hyperactive transposase expression plasmid pCMV-SB11 (465), the mutant transposase expression plasmid pCMV-SBΔDDE (466), and the Sleeping Beauty transposon plasmid pT2/HB (467) were kind gifts from Scott McIvor at the University of Minnesota.

3.1.3. Cell Lines

293T	Human embryonic kidney cell line (HEK293T)
HeLa	Human cervix epitheloid carcinoma cell line

3.1.4. Antibodies

MAB2798	Mouse anti-SB11 monoclonal	R&D Systems
SC8334	Rabbit anti-eGFP polyclonal	Santa Cruz Biotechnology
NA934V	Donkey anti-rabbit HRP-linked	ECL
HAF018	Donkey anti-mouse HRP-linked	R&D Systems

3.1.5. PCR

Primers	Invitrogen
GoTaq DNA Polymerase	Promega
Pfu DNA Polymerase	Promega

3.1.6. Quantitative Real-Time PCR

Primers	Invitrogen

Probes	MWG
Platinum qPCR SuperMix-UDG with ROX	Invitrogen

3.1.7. Kits

Topo-TA Cloning Kit	Invitrogen
Plasmid preparation	Qiagen
QiaQuik Gel Extraction	Qiagen
QiaQuik PCR Purification	Qiagen
HIV-1 p24 ELISA	Retrotek

3.1.8. Western Blotting

Western blotting reagents and equipment were obtained from the Invitrogen NuPAGE system (Sample Buffer, Antioxidant, SeeBlue Plus 2 Pre-stained Standard, NOVEX Bis-Tris Gels, MES Buffer, Transfer Buffer). The chemiluminescent ECL Western Blotting Substrate was supplied by Pierce.

3.1.9. Bacteria

Escherichia coli strains used:

XL1-Blue — Stratagene

Genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI*^{q}*ZΔM15* Tn*10* (Tet^{R})]

Stbl3 — Invitrogen

Genotype: *mcrB mrr hsdS20(*r_B^-, m_B^-*) recA13 supE44 ara-14 galK2 lacY1 proA2 rpsL20(*Str^{R}) *xyl-5* λ^- *leu mtl-1* F$^-$

Genes listed signify mutant alleles, except for genes on the F' episome which are wild-type unless indicated otherwise.

3.2. Methods

3.2.1. Growth and Maintenance of E. coli

Escherichia coli (*E. coli*) were grown in LB broth (1% bactotryptone, 0.5% yeast extract, 1% NaCl) with shaking at 250rpm or streaked out on solid LB agar plates (1.5% agar dissolved in LB by heating) and maintained at 37°C. LB broth or agar containing 100µg/ml ampicillin was used for antibiotic selection. Bacterial cultures containing plasmids were stored at -80°C in LB broth containing 15% (v/v) glycerol.

3.2.2. Production of Electrocompetent XL-1 Blue E. coli

1L LB broth was inoculated 1/100 with an overnight culture of XL-1 Blue *E. coli* and shaken at 37°C until the culture reached an OD600 of 0.6. The culture was chilled on ice for 15min and centrifuged at 6000*g* for 10 minutes at 4°C. The pellet was resuspended in 10% (v/v) glycerol to a volume equal to the original culture and centrifuged as before. The pellet was resuspended in 10% glycerol to 1/2 volume of the original culture and centrifuged as before. The pellet was resuspended in 10% glycerol to 1/4 volume of the original culture and centrifuged as before. The pellet was resuspended in 10% glycerol to 1/50 volume of the original culture and centrifuged as before. The pellet was resuspended in 10% glycerol to 1/500 volume of the original culture and 50µl aliquots were stored at -80°C.

3.2.3. Production of Chemically Competent Stbl3 E. coli

The method of Inoue (468) was used to produce high efficiency chemically competent *E. coli.* A 5ml starter culture of LB broth was made by inoculation with a single colony of Stbl3 *E. coli* and shaken at 37°C overnight. The next day this was used to inoculate 250ml SOB (2% bactotryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl autoclaved and 10mM $MgCl_2$ and 10mM $MgSO_4$ sterile-filtered and added before use) and culture was shaken at 18°C to an OD600 of 0.6. The culture was chilled on ice for 15min and centrifuged at 2500*g* for 5min at 4°C. The pellet was gently resuspended in 80ml TB (10mM PIPES, 15mM $CaCl_2$, 250mM KCl, 55mM $MnCl_2$, 1M KOH to pH6.7), chilled on ice for 10min, and centrifuged as before. The

pellet was resuspended in 20ml 7% DMSO in TB, chilled on ice for 10min, and 50µl aliquots were snap frozen in liquid nitrogen and stored at -80°C.

3.2.4. Transformation of Electrocompetent E. coli

10pg plasmid or 200ng desalted ligation products were gently mixed with a 50µl aliquot of electrocompetent *E. coli*, transferred to a 2mm electroporation cuvette, and electroporated at 2.5kV 200Ω 25µF with a Bio-Rad Gene Pulser II. 1ml LB broth was added to the cuvette and the combined contents were removed into a new tube and shaken at 37°C for 1 hour before plating out on an LB agar plate.

3.2.5. Transformation of Chemically Competent E. coli

10pg plasmid or 200ng ligation products were gently mixed with a 50µl aliquot of electrocompetent *E. coli* and stored on ice for 30min. The cells were heated to 42°C for 45s and chilled on ice for 2min. The cells were transferred to a new tube containing 250µl SOC medium (20mM glucose in SOB as described in Section 3.2.3) and shaken at 37°C for 1 hour before plating out on an LB agar plate.

3.2.6. Plasmid DNA Preparation

E. coli containing plasmids were cultured by shaking in LB broth with the relevant antibiotic at 37°C overnight. Plasmid DNA was extracted by alkaline lysis using a Qiagen Spin Miniprep (≤20µg plasmid DNA), Qiagen Plasmid Maxiprep (≤500µg), or Qiagen Spin Megaprep (≤2.5mg) kit according to the manufacturer's instructions. DNA concentration was quantified by measuring the ABS_{260} on a NanoDrop ND-1000 spectrophotometer.

3.2.7. Ethanol Precipitation of DNA

DNA in solution was precipitated by adding 1/10 volume of 3M sodium acetate pH5.2 and 2 volumes of 100% ethanol, storing at -20°C for 1 hour, and centrifuging at 13,000rpm in a microcentrifuge for 30min. The pellet was washed with 70% ethanol, centrifuged as before, and resuspended in distilled water or TE (10mM Tris-HCl pH 8.0, 1mM EDTA).

3.2.8. Restriction Endonuclease Digestion of Plasmid DNA

1-2µg plasmid was digested with one or two restriction endonucleases (<10% final volume), 1X buffer (provided by the enzyme manufacturer), and 0.1mg/ml BSA in a final volume of 20µl. Digestion was performed for 1-2 hours at the appropriate temperature for the enzyme being used.

3.2.9. Filling of 5' Single Stranded DNA Overhangs

5' overhangs in 1µg digested DNA were filled using the 5'-3' polymerase activity of 1U DNA Polymerase I (Klenow Large Fragment) in 1X DNA Polymerase I buffer (provided by the enzyme manufacturer), 0.1mg/ml BSA, and 40µM of all four dNTPs in 20µl at room temperature for 10min.

3.2.10. Exonuclease Removal of 3' Single Stranded DNA Overhangs

3' overhangs in 1µg digested DNA were removed using the 3'-5' exonuclease activity of 5U T4 DNA Polymerase in 1X T4 DNA Polymerase buffer (provided by the enzyme manufacturer), 0.1mg/ml BSA, and 100µM of all four dNTPs in a total volume of 20µl at 37°C for 5min.

3.2.11. Dephosphorylation of 5' DNA Termini

5' phosphate groups were removed from up to 10pmol digested DNA ends using 0.05U Calf Intestinal Alkaline Phosphatase in 1X CIAP buffer (provided by the enzyme manufacturer) in a total volume of 50µl at 37°C for 1 hour. Dephosphorylation of recessed 5' phosphate groups was enhanced by alternating incubations at 37°C and 56°C every 15 minutes for 1 hour and adding fresh enzyme after 30 minutes.

3.2.12. Agarose Gel Electrophoresis

0.5-2% high melting point agarose or 4% low melting point agarose was dissolved in 1X TAE buffer (40mM Tris-acetate, 5mM EDTA) by heating and 0.5µg/ml ethidium bromide was added before allowing the gel to set within an appropriate form. DNA samples were loaded with Orange G loading dye (10% glycerol, 0.1% Orange G), a

1kb Plus ladder was added as a standard and fragments were separated by electrophoresis in 1X TAE at 80-120V and visualised by exposure to ultraviolet light using a UviDoc gel documentation system.

3.2.13. Agarose Gel Purification of DNA Fragments

Fragments separated by gel electrophoresis were excised from the gel using a scalpel and purified using the Qiaquick Gel Extraction kit. This uses a chaotropic salt solution to dissolve the agarose gel matrix and reversible DNA binding to a silica filter to purify the DNA.

3.2.14. Ligation of DNA Fragments

100-200ng digested plasmid backbone was ligated to insert DNA in a 1:1 or 1:3 molar ratio using 3U T4 DNA ligase and 1X T4 DNA ligase buffer (provided by the enzyme manufacturer) in a total volume of 20µl. Ligation of sticky ends was performed for 3 hours at room temperature and ligation of blunt ends was performed at 15°C overnight. Ligation products were subsequently transformed into competent *E. coli* to recover successful events.

3.2.15. Topoisomerase-T/A Cloning

4µl fresh Taq polymerase-amplified PCR products were cloned into 10ng pCR2.1-TOPO vector in the presence of 200mM NaCl 10mM $MgCl_2$ at room temperature for 30min. Products were immediately transformed into competent *E. coli* to recover successful events.

3.2.16. Isolation of Mammalian Genomic DNA by Salting Out

Episomal and chromosomal DNA was extracted from cells using the salting out method of Miller (469). 10^6 cells were suspended in 3ml nuclear lysis buffer (10mM Tris-HCl pH8.2, 0.4M NaCl, 2mM Na_2EDTA) to which 1/15 volume 10% SDS and 1/6 volume Proteinase K solution (2mg/ml Proteinase K, 1% SDS, 2mM Na_2EDTA) were added to lyse the cells. Lysis was performed at 37°C overnight. Cell debris was precipitated by the addition of 1/4 volume 6M NaCl followed by vigorous shaking. The debris was pelleted by centrifugation at 2500rpm for 15min. The supernatant was

removed into a fresh tube and the DNA was precipitated by the addition of 2 volumes of ethanol. The DNA was pelleted at 4000rpm for 30min, washed with 70% ethanol, and dissolved in TE at 37°C overnight.

3.2.17. Southern Blotting

10µg genomic DNA was digested with 10-40U of restriction enzymes at 37°C overnight. Digestion products were separated by electrophoresis on a 0.8% agarose gel. The gel was shaken in 0.25M HCl for 30min, rinsed in distilled water, and then shaken in denaturation solution (1.5M NaCl, 0.5M NaOH) for two periods of 20min (replacing the solution between periods). The gel was rinsed in distilled water and shaken in neutralisation solution (1.5M NaCl, 0.5M Tris pH7) for two periods of 20min (replacing the solution between periods). The DNA was then blotted overnight onto an Amersham Hybond-N membrane (GE Life Sciences, Piscataway, NJ) by capillary transfer using 20X SSC (3M NaCl, 0.3M trisodium citrate dihydrate). The membrane was washed in 2X SSC and dried.

To visualise bands containing the eGFP coding sequence, a radiolabelled probe was produced by restriction digestion of plasmid pLNT/SFFV-eGFP-WPRE plasmid with BamHI and XbaI followed by gel extraction of the 750bp eGFP fragment. This was labelled using the Megaprime DNA Labelling System (GE Life Sciences, Piscataway, NJ). Briefly, the fragment was denatured by heating to 95°C and labelling buffer containing random 6nt primers, Klenow large fragment DNA polymerase, and 20µCi ^{32}P dCTP were added. The DNA polymerisation reaction was incubated at 37°C for 10min. Unincorporated radioisotope was then removed by centrifugation through an Amersham MicroSpin-s-300 HR Column (GE Life Sciences, Piscataway, NJ), eluting radiolabelled probe.

The membrane was wet in 2X SSC and placed in a hybridisation tube. The membrane was covered with Church mix (1% BSA, 7% SDS, 0.4M Na_2HPO_4, 0.1M NaH_2PO_4) and pre-hybridised by rotating at 68°C for 1h. The probe was denatured by heating and added to the hybridisation tube. Hybridisation was performed by rotating at 68°C overnight. The next day, the membrane was washed at 65°C with 0.5% SDS and decreasing concentrations of SSC (2X, 0.5X, 0.2X, 0.1X) until the background signal from the membrane as determined with a Geiger counter on an irrelevant region of the

membrane was low. The activity was visualised by exposing a phosphor screen to the membrane overnight and imaging using a Molecular Dynamics Typhoon 9410 phosphorimager (GE Life Sciences, Piscataway, NJ).

3.2.18. PCR Amplification of Fragments for Plasmid Subcloning

PCR was performed using 100ng template DNA, 1μM of each primer, 200μM each dNTP, 1.25U Pfu polymerase, and 1X Pfu polymerase buffer (provided by the enzyme manufacturer) in a total volume of 50μl. Initial denaturation was performed at 95°C for 2min and followed by 30 amplification cycles consisting of denaturation at 95°C for 30s, primer annealing at 58°C for 30s (the annealing temperature was adjusted with reference to the primer melting temperatures), extension at 72°C for 1min/kb, and a final extension step of 5min at 72°C.

3.2.19. Screening Transformed E coli by Colony PCR

Bacterial colonies were picked directly into wells in a 96-well PCR plate containing 1μM of each primer, 2mM $MgCl_2$, 200μM each dNTP, and 1X GoTaq polymerase buffer (provided by the enzyme manufacturer) in a total volume of 50μl. The block was heated to 100°C for 5min to lyse the cells. 1.25U GoTaq polymerase was added to each well before standard PCR temperature cycling.

3.2.20. Quantitative Real-Time PCR (qPCR)

Approximately 100ng of genomic DNA was used as a template for each reaction. Reactions were performed in triplicate using 0.9 μM of each primer, 0.2 μM of fluorescent probe, and the Platinum qPCR SuperMix-UDG with ROX mastermix (Invitrogen). Real time PCR was carried out for 40 cycles of [95 °C 15s; 60°C 1min] and quantified using an ABI Prism 7000 (Applied Biosystems, Foster City, CA). Plasmid standards containing the human β-actin coding sequence or HIV-1 vector sequences were a kind gift from Mike Blundell at the Institute of Child Health, University College London.

Human β-actin qPCR

Forward primer 5′-TCA CCC ACA CTG TGC CCA TCT ACG A-3′

Reverse primer	5′-CAG CGG AAC CGC TCA TTG CCA ATG G-3′
Probe	5′-FAM-ATG CCC TCC CCC ATG CCA TCC TGC GT-TAMRA-3′

WPRE qPCR

Forward primer	5’-TTC TCC TCC TTG TAT AAA TCC TGG TT-3’
Reverse primer	5’-CGC CAC GTT GCC TGA CA-3’
Probe	5’-FAM-CTG TCT CTT TAT GAG GAG TTG TGG CCC G-TAMRA-3’

HIV-1 primer binding site qPCR

Forward primer	5'-TGT GTG CCC GTC TGT TGT GT-3’
Reverse primer	5'-GAG TCC TGC GTC GAG AGA GC-3’
Probe	5'-FAM-CAG TGG CGC CCG AAC AGG GA-TAMRA-3'

3.2.21. Ligation-Mediated PCR (LM-PCR) for Recovery of Integration Sites

LM-PCR linker and primers

Linker+	5’-GTA ATA CGA CTC ACT ATA GGG CTC CGC TTA AGG GAC CGC ATG-3’
Linker-	5’-P-CGG TCC CTT AAG CGG AG-3’ (note 5’ phosphate)
Linker first round	5’-GTA ATA CGA CTC ACT ATA GGG C-3’
Linker nested	5’-AGG GCT CCG CTT AAG GGA C-3’
3’ LTR first round	5’-AGT GCT TCA AGT AGT GTG TGC C-3’
3’ LTR nested	5’-GTC TGT TGT GTG ACT CTG GTA AC-3’

IRDRR first round 5’-ACC CAC TGG GAA TGT GAT GAA AG-3’

IRDRR nested 5’-AAT AAA GTG GTG ATC CTA ACT GAC C-3’

All nucleotides were obtained from Invitrogen at 25nmol scale and desalted purity. Linker oligos were used in all LM-PCR reactions, 3' LTR oligos were used to recover HIV-1 integration sites and IRDRR oligos were used to recover SB integration sites. Sequences were modelled after Wu *et al* (470) and Ikeda *et al* (471).

Linker cassette was produced by heating 100pmol Linker+ and 100pmol Linker- in 200µl 250mM Tris pH7.5 100mM $MgCl_2$ to 95°C for 5min in a hot block and allowing the solution to cool slowly to room temperature. Cassette was aliquoted and stored at -20°C.

1µg genomic DNA was digested with NlaIII overnight and subsequently with BamHI to prevent recovery of any remaining transposon-plasmid or transposon-lentivector junctions. The mixture was heat inactivated at 65°C for 20min. The digested DNA was ligated to 1µl of cassette at room temperature for 3 hours and heat inactivated at 70°C for 10min. First round PCR was performed using one tenth of the ligation reaction and GoTaq polymerase under conditions 95°C 2min; 30x [95°C 30s; 55°C 30s; 72°C 1min]; 72°C 5min. 1/100 of the product was used for the nested PCR reaction under the same cycling conditions.

In the first round of integration site sequencing, these LM-PCR products were Topo-TA cloned into the pCR2.1 vector, transformed into Stbl3 cells and the plasmids were sequenced using conventional Sanger sequencing by Functional Biosciences of Madison, WI.

In the second round of integration site sequencing, LM-PCR products were sent for 454 pyrosequencing (472) by our collaborator Richard Gabriel in the laboratory of Manfred Schmidt at the German Cancer Research Center (DKFZ) in Heidelberg, Germany. In this method, additional PCR was performed on LM-PCR products using fusion primers A and B to add adaptor sequences for pyrosequencing. Each fusion primer A contained an individual barcode sequence of 6 nucleotides (N) that allowed identification of different samples sequenced in the same sequencing run. Briefly,

primer A was joined to a lentiviral vector specific LTR primer (fusion primer-A-LTR: Sequence: 5´-GCC TCC CTC GCG CCA TCA GNN NNN NTG TGT GAC TCT GGT AAC TAG) or Sleeping Beauty IRDR-R specific primer (fusion primer A-IRDRR, Sequence: 5´-GCC TCC CTC GCG CCA TCA GNN NNN NGT ATT TGG CTA AGG TGT ATG) and fusion primer B was joined to a linker cassette specific primer (fusion primer B-LK, Sequence: 5´-GCC TTG CCA GCC CGC TCA GAG GGC TCC GCT TAA GGG AC). 40ng of purified LM-PCR products were amplified by fusion primer PCR under conditions 95°C 2min; 12x [95°C 45s; 60°C 45s; 72°C 60s]; 72°C 5min. A sample of the PCR products was analysed by agarose gel electrophoresis before pyrosequencing using a 454 GenomeSequencer Flx (454 Life Sciences, Branford, CT).

3.2.22. Integration Site Bioinformatics

Raw sequence reads obtained after sequencing were trimmed and aligned to the human genome. Integration sites were considered to be valid if a vector-genome junction sequence was present and the flanking genomic region had a unique sequence match of at least 95% after alignment to the human genome release hg18 (University of California at Santa Cruz, (UCSC) RefSeq genes and RepeatMasker (473). Chromosome graphs were generated using the UCSC Genome Graphs tool. HeLa cell expression data were obtained from the public Gene Expression Omnibus database with accession number GSM157868, a dataset obtained from total RNA using a GeneChip Scanner 3000 (Affymetrix). Sequence logos were produced with the WebLogo tool (474).

3.2.23. Propagation and Storage of Mammalian Cell Lines

Adherent cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) containing GlutaMAX supplemented with 10% foetal calf serum (FCS) and 1% penicillin/streptomycin (hereafter referred to as 'complete DMEM'). Suspension cells were maintained in RPMI 1640 medium containing GlutaMAX, serum, and penicillin/streptomycin as before. All cells were incubated at 37°C and a 5% CO_2 atmosphere. For long term storage, 1-5 x 10^6 cells were suspended in freezing medium (50% complete DMEM, 40% FCS, 10% DMSO) and frozen to -80°C slowly overnight in an isopropanol freezing box before long term storage in liquid nitrogen.

3.2.24. G418 Selection of HeLa Cell Colonies

HeLa cells were transfected or transduced in 24-well plates with a neomycin resistance cassette and incubated without selection for three days. On the third day cells were trypsinised and $1/10^{th}$ or $1/50^{th}$ of the suspension was seeded into a fresh 24-well plate. Cells were maintained without trypsinisation in complete DMEM containing 1mg/ml G418 for two weeks with a medium change every three days.

3.2.25. Quantification of Mammalian Cell Colonies

Colonies were fixed with 1% paraformaldehyde in PBS and stained with crystal violet in PBS. Plates were air-dried overnight and scanned using a Hewlett-Packard ScanJet scanner. The number of colonies in each image file was counted using CellProfiler image-processing software (475).

3.2.26. Lentiviral Vector Preparation

1.2×10^7 HEK293T cells were seeded in $175cm^2$ flasks one day before transfection to reach >90% confluence. For each flask, 50µg vector plasmid, 32.5µg packaging plasmid pCMV-dR8.74 (for integrating virus) or pCMV-dR8.74 D64V (non-integrating virus), and 17.5µg vesicular stomatitis virus envelope plasmid pMDG2 was added to 5ml Optimem and 0.22µm filtered. 1µl PEI was added to 5ml Optimem and also 0.22µm filtered. The two mixtures were combined and allowed to complex for 20min. Cells were washed with PBS, overlaid with the complex, and incubated at 37°C 5%CO_2 for 4 hours. The complex was removed and replaced with complete DMEM. 24 hours later, the medium was replaced. 48 and 72 hours after transfection the medium was removed, 0.22µm filtered, and centrifuged at 98,000*g* for 2 hours. Virus pellets were resuspended in Optimem and aliquots were stored at -80°C.

3.2.27. Titration of Lentiviral Vector Preparations

3.2.27.1. Expression Titre

10^5 cells of the cell line of interest were seeded into each well of a 24-well plate one day before transduction. For transduction, a 5-fold dilution series of virus in 500µl aliquots of complete medium was performed. Each aliquot of diluted virus was used

to replace the medium in one well. Transgene expression was measured by flow cytometry 48 hours post-transduction or by G418-resistant colony counting as described in Section 3.2.24. Expression titre was calculated by selecting the well in which 5-15% of transduced cells expressed the transgene of interest and dividing the number of transduced cells in this well by the volume of virus used to transduce them.

3.2.27.2. Vector DNA Copy Number Titre

Cells were transduced as above. 48 hours post-transduction, cells were trypsinised and DNA was extracted as described in Section 3.2.16. The WPRE copy number and the total cell number were determined by quantitative PCR as described in Section 3.2.20. The titre was calculated as the total WPRE copy number in all cells in the well divided by the volume of virus used to transduce them.

3.2.27.3. Physical Titre

The mass of HIV-1 p24 antigen in the lentiviral vector preparation was determined using the Retrotek HIV-1 p24 Antigen ELISA kit according to the manufacturer's instructions. Concentrated vector preparation was diluted 1:90 in 450µl Optimem before the initial lysis step. Proteins present in the lysed vector were adhered to the wells of a microplate, and biotinylated anti-p24 antibody allowed to bind selectively to p24 antigen. Streptavidin-horseradish peroxidase was bound to the biotinylated antibody. The oxidation of tetramethybenzidine by peroxidase was quantified by measurement of A_{450} in a plate reader. The concentration of p24 antigen in the preparation was estimated from an A_{450} – p24 concentration standard curve using known standards supplied with the kit.

3.2.27.4. RNA genome quantification

Virion particles in concentrated lentiviral supernatant were lysed using TRIzol reagent for 5min and extracted with chloroform by shaking and centrifugation. The aqueous phase was transferred to a fresh tube, precipitated with isopropanol and pelleted by centrifugation. The pellet was washed with 75% ethanol and dissolved in RNase-free water. Plasmid carryover in the preparation was destroyed by digesting with 1U amplification grade DNase I per µg of RNA at room temperature for 15min. The

DNase I was heat-inactivated at 65°C for 10min. The RNA was reverse-transcribed using M-MLV reverse transcriptase and random primers according to the manufacturer's instructions. The resultant cDNA was used as a template for quantitative PCR analysis.

3.2.28. Transduction of Target Cells

10^5 cells of the cell line of interest were seeded into each well of a 24-well plate one day before transduction. Transductions were performed by adding virus to 500µl of complete medium in each well.

3.2.29. Transfection of Target Cells

10^5 cells of the cell line of interest were seeded into each well of a 24-well plate one day before transduction. Plasmid DNA was complexed with Lipofectamine 2000 in a 1:2 ratio in Optimem for 20min. Complex was added directly to 500µl complete medium in each well.

3.2.30. Flow Cytometry

The fluorescence of cells expressing eGFP was measured using a CyAn ADP flow cytometry analyser (Dako, Glostrup, Denmark). Data was analysed using FlowJo version 7.2.4 (Tree Star Inc, Ashland, OR).

3.2.31. Western Blotting

Cells were lysed in lysis buffer (1% Nonidet P40, 130mM NaCl, 20mM Tris-HCl pH 8.0, 10mM NaF, 2mM sodium orthovanadate, 1% aprotinin, 10µg/ml leupeptin, 1mM phenylmethylsulfonyl fluoride, 2mM EDTA, 2µM Cytochalasin D, 2µM Latrunculin B) at 95°C for 2 minutes. 1X NuPAGE LDS Sample Buffer and 1X NuPAGE Sample Reducing Agent were added to the lysis products. The samples and the SeeBlue Plus2 Prestained Standard were separated by electrophoresis in a Novex 4-12% Bis-Tris gel in 1X MES buffer at 200V for 35min. The separated proteins were wet-transferred onto a methanol-activated Immobilon-P PVDF membrane using the X-Cell II Blot Module and 1X NuPAGE Transfer Buffer containing 0.1% NuPAGE Antioxidant and 10% methanol at 30V for 1 hour. The membrane was blocked in PBST (PBS

containing 0.5% Tween-20 and 5% BSA) for 1 hour. Primary antibody staining was performed for 1 hour with a 1:200 dilution of Santa Cruz rabbit anti-eGFP or a 1:500 dilution of R&D Systems mouse anti-SB11 in PBST. After washing three times with PBST, secondary antibody staining was performed for 45min with either a 1:2000 dilution of ECL horseradish peroxidase conjugated donkey anti-rabbit or a 1:1000 dilution of R&D Systems horseradish peroxidase-conjugated donkey anti-mouse in PBST. After washing three times with PBST, peroxidase activity was visualised using the Pierce ECL Western Blotting Substrate and a UviChemi chemiluminescence detection system.

3.2.32. Confocal Microscopy

Cells were seeded onto poly-L-lysine slides and transfected or transduced in complete DMEM culture medium. 48 hours later, cells were washed with PBS, fixed with 1% paraformaldehyde, and stained with a 1:200 dilution of DiI (Invitrogen) or a 1:40 dilution of rhodamine phalloidin (Invitrogen) in PBS. Stained cells were washed in PBS and mounted with Vectashield Mounting Medium with DAPI (Vector Laboratories, Burlingame, CA). Confocal images were obtained using a Leica SP2 confocal laser scanning microscope (Leica, Milton Keynes, UK).

4. Lentivirus-Transposon Hybrid Vectors Expressing eGFP

4.1. Aims

- To incorporate the transposon and transposase elements of the Sleeping Beauty system into lentiviral vectors
- To demonstrate expression of the Sleeping Beauty transposase from a non-integrating lentiviral vector
- To demonstrate transposition from a non-integrating lentiviral vector into host chromosomes

4.2. Introduction

Successful clinical trials using gammaretroviral vectors to treat haematopoietic disorders have highlighted the usefulness of integrating vectors in mitotic tissues (476;477), but serious adverse events have been observed in which vector integration near proto-oncogenes led to cell transformation (478;479). Gammaretroviral, lentiviral, and foamy virus vectors are all being developed for gene therapy, and all three vectors integrate preferentially near host genes or regulatory elements (480-482). By contrast, vectors based on the Sleeping Beauty transposon integrate almost randomly with respect to these elements and are potentially safer in this respect (483). Sleeping Beauty vectors are conventionally transfected into target cells as plasmids using non-viral techniques. However, non-viral transfection of important targets such as haematopoietic stem cells is higly inefficient. A hybrid vector combining the efficient cell and nuclear entry properties of HIV-1 vectors with the random integration profile of Sleeping Beauty would represent a theoretically safer alternative compared to conventional lentiviral vectors. This chapter describes the initial construction and testing of this hybrid vector using eGFP expression as a reporter for integration efficiency.

Sleeping Beauty transposition within a target cell requires two components, namely a DNA transposon and the transposase protein to mediate the transposition reaction.

The two components can be delivered to target cells *in cis* on the same vector molecule or *in trans* through simultaneous transduction or transfection with a transposon vector and a transposase-expression vector (484). The *cis* configuration is preferable from a clinical perspective as it requires only one vector to be developed, approved, and administered. There is also a theoretical efficiency gain to be made through guaranteeing the presence of transposon and transposase in the same cells, though this has not been widely reported. However, the need to accurately control the concentration of transposase in target cells is better served by a *trans* configuration which allows the two components to be titrated against one another, and generally the *cis* configuration is used to refine a *trans* Sleeping Beauty system once it is established. The *trans* approach was adopted in this chapter.

Since the aim of the project was to reduce the risk of insertional mutagenesis due to integration mediated by the HIV-1 integrase, Sleeping Beauty components were delivered to target cells using non-integrating lentiviral vectors in which the catalytic activity of HIV-1 integrase is blocked by the D64V mutation (485). The lentivirus-transposon vector was initially constructed so that the transposon was in the reverse orientation relative to the lentiviral backbone (Figure 4.1) in order to prevent termination of transcription at a known polyadenylation signal present within the Sleeping Beauty right inverted repeat (486).

The hybrid vector strategy is illustrated in Figure 2.11. Target cells are co-transduced with two non-integrating lentiviral vectors for delivery of transposase expression and the transposon cargo. Following reverse transcription and nuclear localisation, the transposase protein is expressed and imported into the nucleus. Here transposase binds to the transposon inverted repeats and catalyses excision of the transposon from the episomal lentiviral vector backbone. The resulting preintegration complex is mobile and moves within the nucleus until it encounters a suitable target DNA. Transposase then catalyses the integration of the transposon into the target DNA. The lentiviral episomes are subsequently diluted through repeated cell division.

4.3. Cloning and Testing of Lentivirus-Transposon Vectors

4.3.1. Plasmid Subcloning of eGFP Transposon and Transposase Expression Vectors

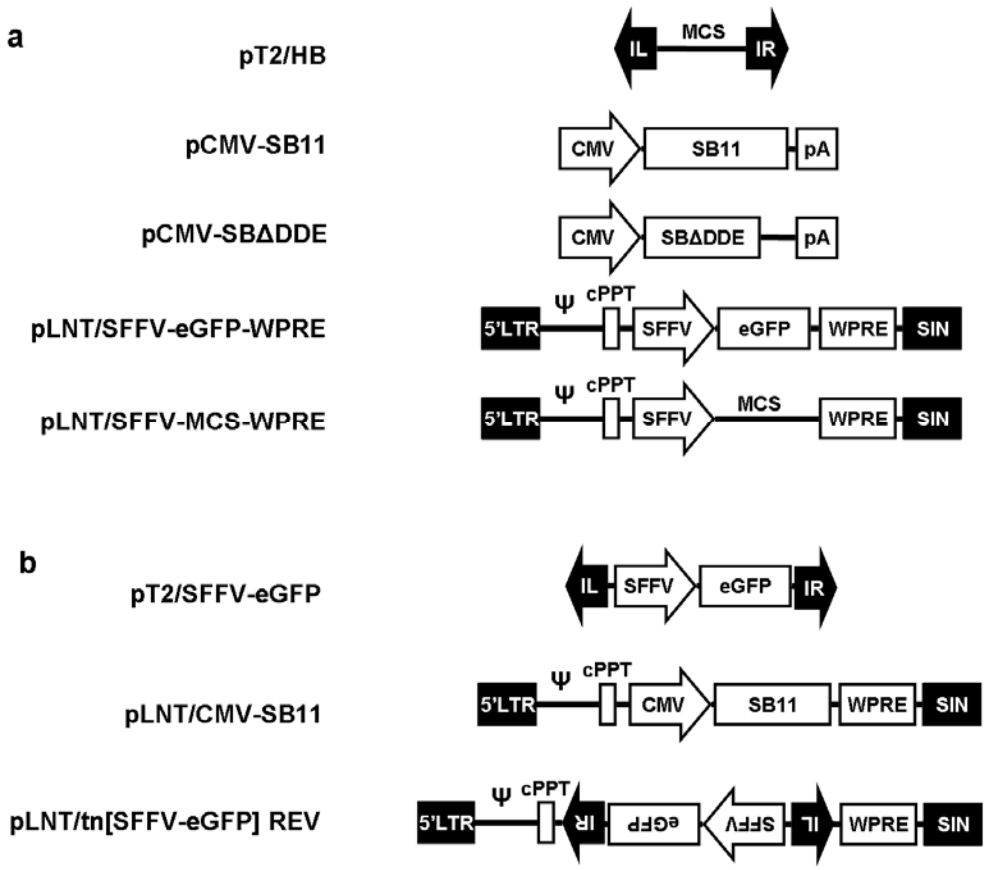

Figure 4.1 Vector map of the eGFP transposon and transposase expression plasmids.
(a) parental plasmids; **(b)** plasmids cloned in this study. MCS, multiple cloning site; SBΔDDE, truncated Sleeping Beauty transposase in which the DDE catalytic domain has been removed; pA, polyadenylation signal; 5'LTR, HIV-1 5' long terminal repeat; Ψ, HIV-1 RNA packaging signal; SIN, self-inactivating (U3-deleted) HIV-1 long terminal repeat; cPPT, central polypurine tract; CMV, immediate early promoter of human cytomegalovirus; WPRE, woodchuck posttranscriptional regulatory element; IL, Sleeping Beauty transposon left inverted repeat; IR, Sleeping Beauty transposon right inverted repeat; SFFV, spleen focus forming virus U3 promoter; eGFP, enhanced green fluorescent protein.

pT2/SFFV-eGFP

The eGFP expression cassette was derived from lentiviral backbone pLNT/SFFV-eGFP-WPRE, a previously described variant of pHR containing a self-inactivating 3' LTR, a central polypurine tract (cPPT), the spleen focus forming virus U3 promoter (SFFV), an enhanced green fluorescent protein reporter gene (eGFP), and the woodchuck posttranscriptional regulatory element (WPRE) (487). This was digested with BamHI, the ends filled in using DNA polymerase I Large (Klenow) Fragment, and then religated using T4 DNA Ligase to destroy this site. This plasmid was recovered and digested with EcoRI and XbaI to yield a fragment containing SFFV-eGFP. This was cloned into Sleeping Beauty transposon plasmid pT2/HB (488) (a kind gift from Scott McIvor, University of Minnesota) between the EcoRI and XbaI sites to give pT2/SFFV-eGFP.

pLNT/CMV-SB11

The SFFV promoter was removed from pLNT/SFFV-MCS-WPRE by digestion with EcoRI and XhoI. The CMV promoter was obtained from the hyperactive transposase expression plasmid pCMV-SB11 (489) (a gift from Scott McIvor) with EcoRI and XhoI. Ligation of the two plasmids yielded pLNT/CMV-MCS-WPRE.

pCMV-SB11 was digested with SacII to obtain the SB11 transposase coding sequence. This was cloned into pBluescript at the SacII site so that the pBS XhoI site was 5' of SB11 and the pBS SacI site was 3'. The new plasmid was named pBS/SB11.

pLNT/CMV-MCS-WPRE was digested with KpnI, the 3' overhang removed with T4 DNA polymerase, and then digested with XhoI. pBS/SB11 was digested with SacI, blunted with T4 DNA polymerase, and then digested with XhoI to obtain SB11. Ligation of SB11 into the lentiviral backbone yielded pLNT/CMV-SB11.

pLNT/tn[SFFV-eGFP] REV

pT2/SFFV-eGFP was digested with EcoRI. The ends were filled in using DNA polymerase I Large (Klenow) Fragment and then religated using T4 DNA Ligase to destroy the EcoRI site. The transposon from EcoRI-deleted pT2/SFFV-eGFP was

PCR-amplified using primer 5'-TGC CAA GAA TTC GGA TCC CTA TAC AG-3' (this acts as both forward and reverse primer) to give it EcoRI ends. This fragment and lentiviral backbone pLNT/SFFV-MCS-WPRE (another pHR variant containing a multiple cloning site) were digested with EcoRI and ligated to give pLNT/tn[SFFV-eGFP] REV SFFV. The eGFP transposon was obtained from pT2/SFFV-eGFP by BamHI digestion and ligated into the BamHI-digested pLNT/tn[SFFV-eGFP] REV SFFV backbone to give pLNT/tn[SFFV-eGFP] REV. Colonies were screened to ensure the transposon was in the reverse orientation.

As a technical note, cloning of transposon constructs in Stratagene XL-1 Blue *E. coli* proved to be inefficient due to a very high proportion of rearranged plasmids, perhaps as a result of recombination between the highly homologous IRs. When cloning was performed using low recombination activity Invitrogen Stbl3 *E. coli*, the recovery of the desired constructs was less difficult.

4.3.2. Assaying Transposition from a Plasmid Donor

A chromosomal integration assay was established to determine the level of transposition in target cells. Direct measurement of integration through quantitative PCR has been reported but such assays are not commonly used (490). A more conventional approach is to transfect or transduce a dividing cell population with the vector or vectors, allow the cells to divide repeatedly in order to dilute persisting episomal DNA, and at a suitable endpoint measure the proportion of cells which express the reporter gene. It is assumed that reporter gene expression detected after a sufficient number of cell divisions will be the product of integrated DNA.

An assay of this type was used to assess transposition of the eGFP transposon from plasmid pT2/SFFV-eGFP (Figure 4.2). 293T cells were transfected with the transposon plasmid and an equal mass of plasmid expressing either the SB11 transposase or a Sleeping Beauty transposase from which the catalytic domain had been removed (SBΔDDE). Equal efficiency of the initial level of gene transfer was confirmed by flow cytometry 2 days after transfection. Cells were passaged until 14 days post-transfection in order to dilute unintegrated plasmid and the level of persistent gene expression was measured. The level of stable gene expression was greatly increased in cells transfected with the active rather than the truncated

transposase, demonstrating successful transposition of the pT2/SFFV-eGFP transposon.

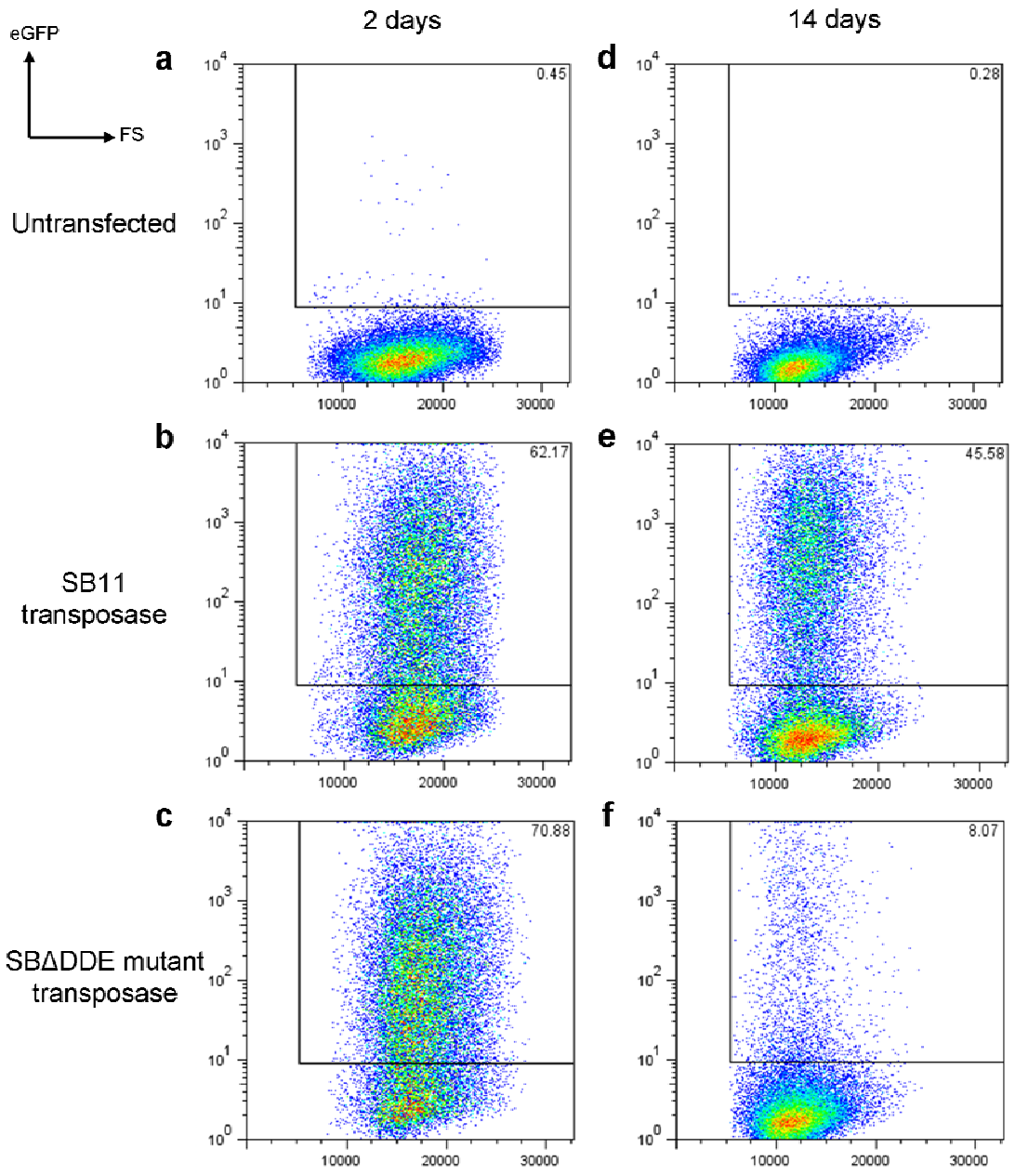

Figure 4.2 Transposition from eGFP transposon plasmid pT2/SFFV-eGFP.

10^5 293T cells were transfected with 400ng pT2/SFFV-eGFP and 400ng of either the active transposase plasmid pCMV-SB11 or the inactive mutant transposase plasmid pCMV-SBΔDDE. eGFP expression was determined by flow cytometry 2 and 14 days post-transfection. The proportion of cells expressing eGFP is given as a percentage within the gated area.

4.3.3. Expression of Transposase from a Non-integrating Lentiviral Vector

Transient transposase expression is desirable in Sleeping Beauty applications in order to minimise the risk of repeated transposon remobilisation. Transgene expression can be lost over time in dividing cells transduced with non-integrating lentiviral vectors through dilution of the episomal lentiviral DNA (491). In order to demonstrate this in the context of transposase expression, integrating and non-integrating lentiviral vectors for transposase expression were produced in parallel using vector plasmid pLNT/CMV-SB11. HeLa cells were transduced with equal volumes of the two vectors and passaged for two weeks. Cells were subcultured every three days, and cell pellets were frozen down to assess transposase expression over time by Western blot (Figure 4.3).

When transposase was expressed from a non-integrating lentiviral vector, expression of the 40kDa SB11 protein was detectable 1 day post-transduction and reached its maximum on day 2. Expression subsequently fell until it was no longer detectable by day 7. Expression from integrating vector was also greatest on day 2 but subsequently remained stable.

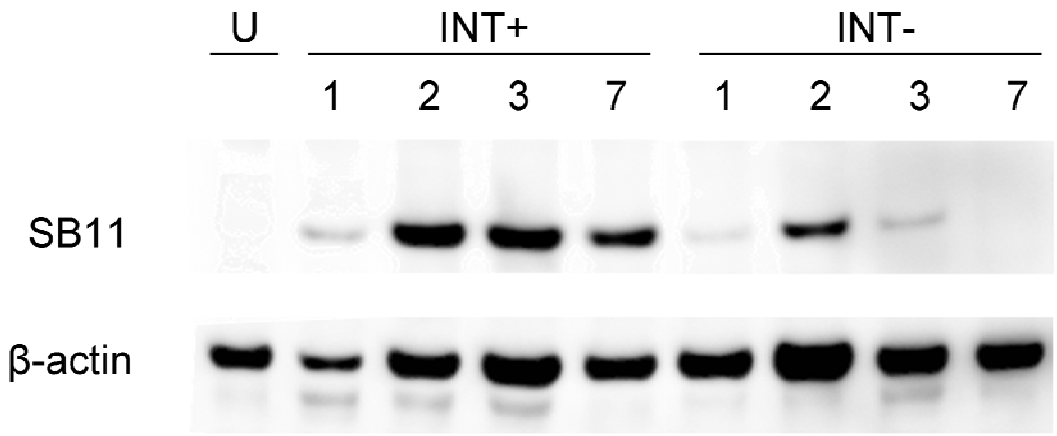

Figure 4.3 Western blot for expression of transposase from lentiviral vectors.

Integrating and non-integrating lentiviral vectors expressing transposase were prepared in parallel using plasmid pLNT/CMV-SB11. 10^6 HeLa cells were transduced with 5μl of virus per well. At 1, 2, 3, and 7 days post-transduction, cell pellets of equal cell number were frozen down for Western blotting. U, untransduced; INT+, integrating vector; INT-, non-integrating vector.

4.3.4. Physical and Functional Titration of Lentiviral Vector Carrying a Transposon in the Reverse Orientation

To test whether lentiviral vectors are functionally able to deliver Sleeping Beauty transposons to target cells, non-integrating lentiviral vector was produced using the pLNT/tn[SFFV-eGFP] REV plasmid. Functional titration was performed by transducing 293T cells with a serial dilution of this vector and assessing eGFP expression by flow cytometry 2 days post-transduction (Figure 4.4).

The functional titre was estimated to be $3.2x10^6$ TU/ml. This is at least two orders of magnitude less than would be expected with a well-packaged lentiviral vector preparation concentrated by ultracentrifugation, and repeated vector preparations did not result in improved titre (data not shown).

To further investigate this problem, the physical titre of the original vector preparation was measured by p24 ELISA. The physical titre measures the concentration of viral proteins in the vector preparation and measures the production of virions by producer cells rather than the functioning of the vector genome in target cells. The physical titre of the original vector preparation was found to be $1.6x10^6$ pg p24/ml while the physical titre of non-integrating LNT/CMV-SB11 was $1.4x10^8$ pg p24/ml. The expected physical titre of a well-packaged lentiviral vector is in the range 10^7-10^8 pg p24/ml after concentration by ultracentrifugation (492), indicating problems in the production of LNT/tn[SFFV-eGFP] REV but not LNT/CMV-SB11. The low physical titre of the LNT/tn[SFFV-eGFP] REV preparation may have been due to cytotoxicity of the vector plasmid to the producer cells, for example caused by contamination during plasmid preparation.

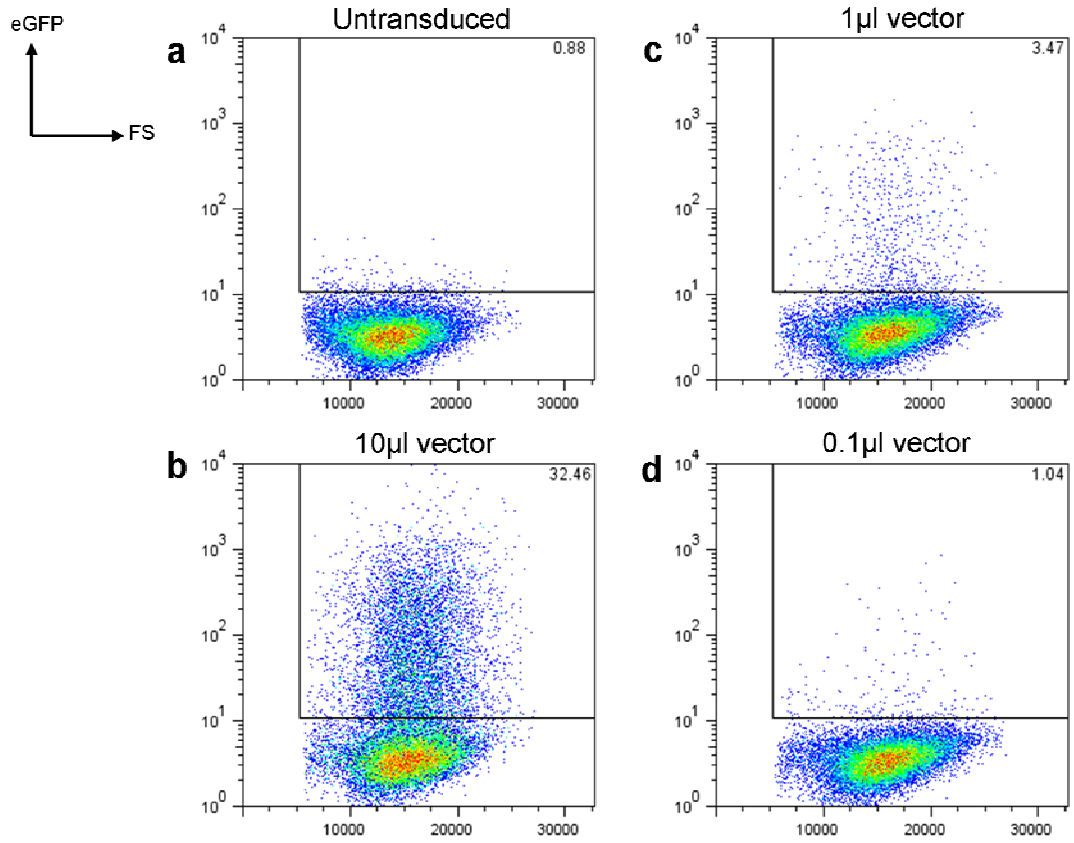

Figure 4.4 Functional titration of eGFP transposon lentivirus LNT/tn[SFFV-eGFP] REV.

10^5 293T cells were transduced with serial dilutions of non-integrating LNT/tn[SFFV-eGFP] REV lentiviral vector. eGFP expression was determined by flow cytometry 2 days post-transduction. **(a)** untransduced; **(b)** 10µl LNT/tn[SFFV-eGFP] REV; **(c)** 1µl; **(d)** 0.1µl. The proportion of cells expressing eGFP is given as a percentage within the gated area.

4.3.5. Quantification of Lentivirus-Transposon Vector Genome Packaging

In order to investigate the low titre of the transposon vector, RNA was extracted directly from concentrated lentiviral preparation, treated with DNaseI to remove contaminating plasmid, and quantified using RT-qPCR (Figure 4.5). Two primer-probe sets were used, one at the 5' end of the vector genome near the primer binding site (PBS), and one at the 3' end of the genome within the WPRE.

The RT-qPCR assay showed no significant difference in copy number between the PBS and WPRE signal for lentivirus-transposon vector. However, when compared to a standard lentiviral vector expressing eGFP, the packaged copy number for both primer-probe sets was over 100-fold lower.

The continuing problem of producing high titre transposon vector preparations raised the possibility that the design of the vector needed to be changed. A relatively straightforward possibility was to reconstruct the transposon vector so that the transposon was in the same orientation as the lentiviral backbone.

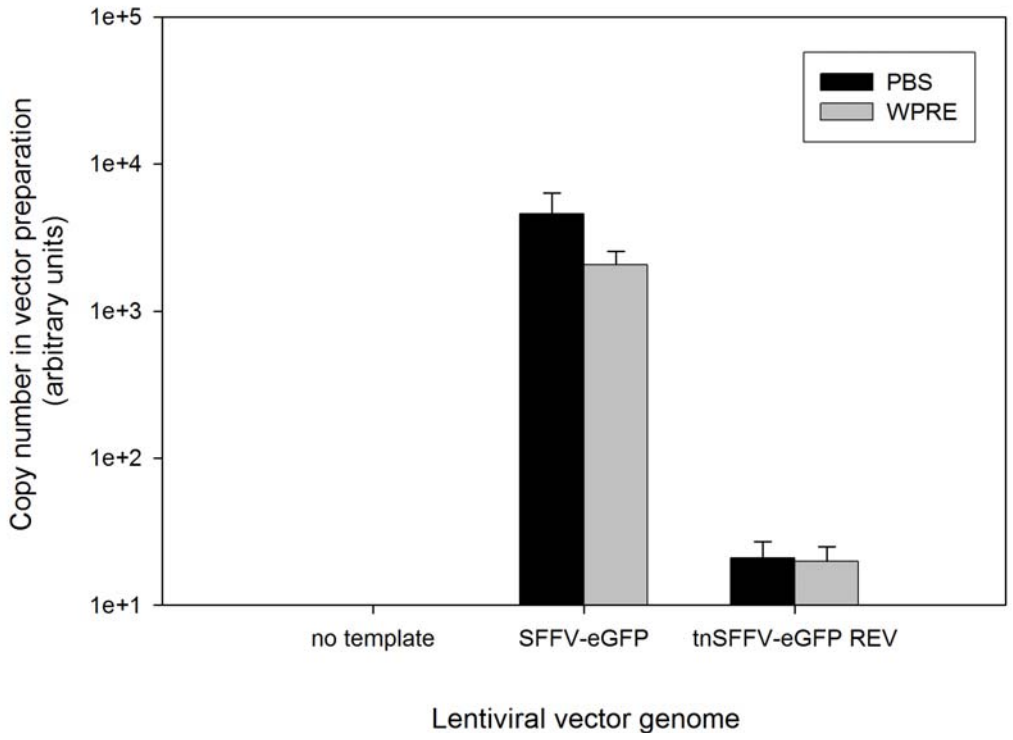

Figure 4.5 Assessment of lentivirus-transposon vector genome packaging and integrity within virions.

RNA was extracted from equal volumes of concentrated lentiviral vector preparation of either LNT/SFFV-eGFP-WPRE or LNT/tn[SFFV-eGFP] REV using Trizol reagent, treated with DNase I, and reverse-transcribed with Moloney murine leukaemia virus reverse transcriptase and random primers. The copy number was determined by quantitative real-time PCR using serially diluted plasmid standards.

4.3.6. Reorientation of the Transposon

The transposon was reorientated so that the HIV-1 LTR promoter and the SFFV promoter were facing in the same direction (Figure 4.6). Although this design makes the polyadenylation signal contained within the right inverted repeat available for vector genomic RNA processing, it was suggested that this signal might be 'leaky' enough to permit production of full length genomic RNA. Plasmid subcloning was performed as described in Section 4.3.1 except that colonies were screened for transposons in the forward orientation.

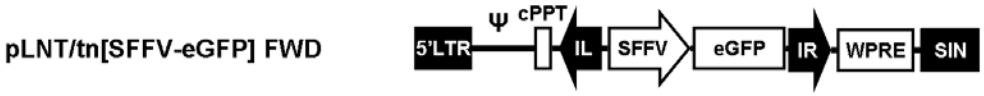

Figure 4.6 Vector map of the forward eGFP transposon.
5'LTR, HIV-1 5' long terminal repeat; Ψ, HIV-1 RNA packaging signal; SIN, self-inactivating (U3-deleted) HIV-1 long terminal repeat; cPPT, central polypurine tract; WPRE, woodchuck posttranscriptional regulatory element; IL, Sleeping Beauty transposon left inverted repeat; IR, Sleeping Beauty transposon right inverted repeat; SFFV, spleen focus forming virus U3 promoter; eGFP, enhanced green fluorescent protein.

To assess the effect of this configuration on vector titre, non-integrating lentiviral vectors were produced in parallel using both pLNT/tn[SFFV-eGFP] FWD and pLNT/tn[SFFV-eGFP] REV. Functional titration was performed by transducing 293T cells with serial dilutions of vector and assessing eGFP expression by flow cytometry 2 days post-transduction (Figure 4.7).

The functional titres were estimated to be 9.9×10^5 TU/ml for LNT/tn[SFFV-eGFP] REV and 3.4×10^8 TU/ml for LNT/tn[SFFV-eGFP] FWD. Subsequent experiments were performed with the FWD vector.

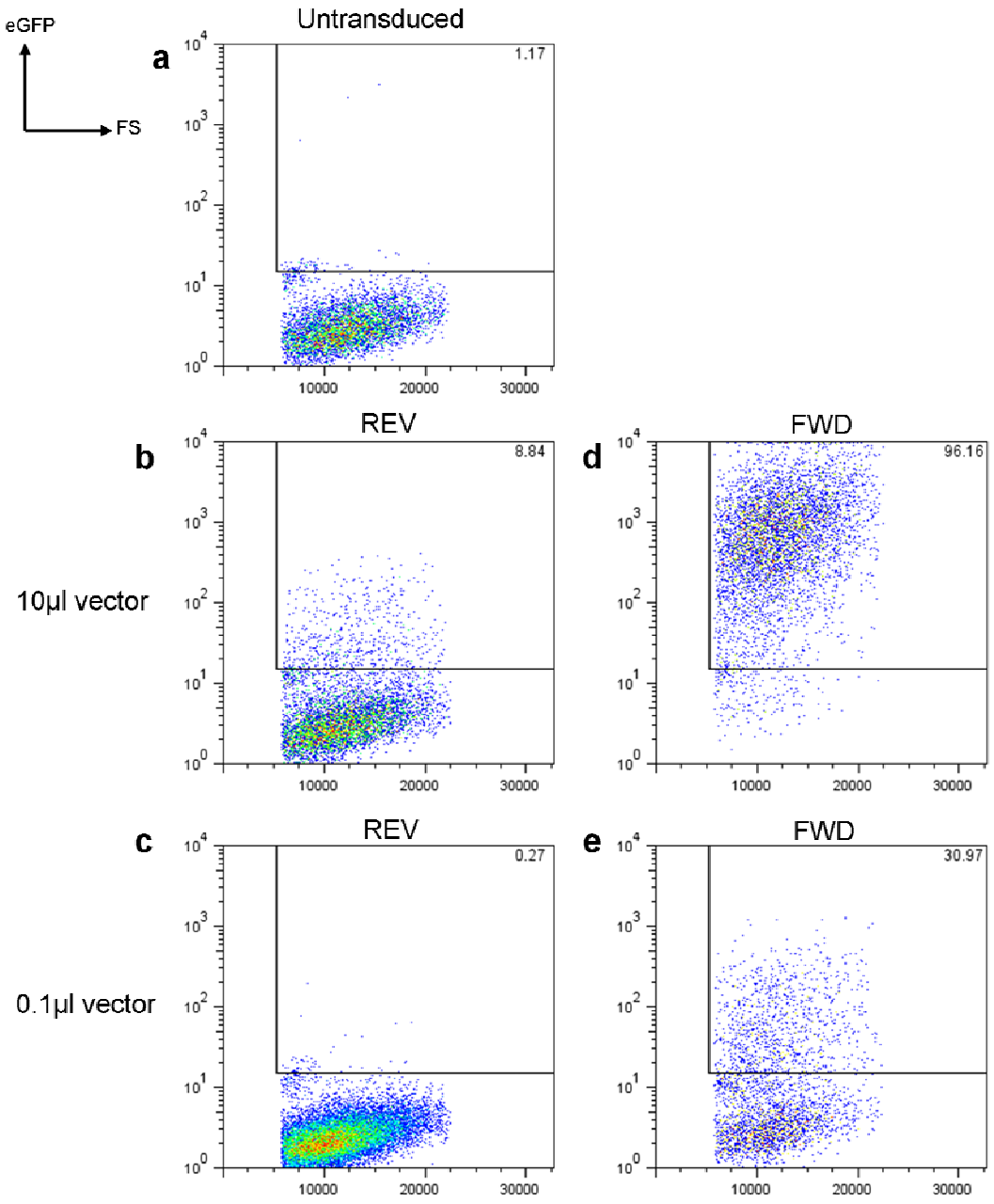

Figure 4.7 The effect of transposon orientation on the functional titre of the lentiviral vector.
10^5 293T cells were transduced with lentiviral vectors in which the transposon was in either the reverse orientation relative to the lentivirus, LNT/tn[SFFV-eGFP] REV, or the forward orientation, LNT/tn[SFFV-eGFP] FWD. eGFP expression was determined by flow cytometry 2 days post-transduction. The proportion of cells expressing eGFP is given as a percentage within the gated area. **(a)** untransduced; **(b)** 10µl LNT/tn[SFFV-eGFP] REV; **(c)** 0.1µl LNT/tn[SFFV-eGFP] REV; **(d)** 10µl LNT/tn[SFFV-eGFP] FWD; (e) 0.1µl LNT/tn[SFFV-eGFP] FWD.

4.3.7. Transposition from Lentiviral Plasmid

The integration assay was repeated in order to demonstrate the ability of plasmid pLNT/tn[SFFV-eGFP] FWD to act as a transposon donor for transposition (Figure 4.8). 293T cells were transfected with pLNT/tn[SFFV-eGFP] FWD and a transposase expression plasmid in a 1:1 ratio as before. Additionally, pLNT/tn[SFFV-eGFP] FWD-transfected cells were transduced with serial dilutions of non-integrating transposase expression lentiviral vector or transfected with serial dilutions of transposase expression plasmid in order to investigate the effects of titrating transposase expression.

When cells were co-transfected with transposon and transposase plasmids in a 1:1 ratio, the proportion of $eGFP^+$ cells after 14 days was approximately 2-fold greater than the background level of expression in cells transfected with transposon plasmid alone. Titration of the transposase expression plasmid showed a decline in reporter gene persistence below 100ng transposase plasmid per 10^5 cells (Figure 4.8f).

When cells were co-transduced with a transposase expression lentiviral vector, the highest proportion of $eGFP^+$ cells observed was 4-fold higher than background and increased with the volume of transposase vector (Figure 4.8e). The proportion of $eGFP^+$ cells was generally higher when transposase expression was provided by lentiviral vector rather than plasmid in this particular experiment, possibly as a result of a suboptimal plasmid transfection.

These results suggest that the pLNT/tn[SFFV-eGFP] FWD plasmid can be used as a transposon donor in the transposition reaction, and that expression of transposase from a non-integrating lentiviral vector can drive the transposition reaction. However, this experiment would need to be repeated with multiple repetitions of each sample for it to have sufficient statistical power to support strong conclusions.

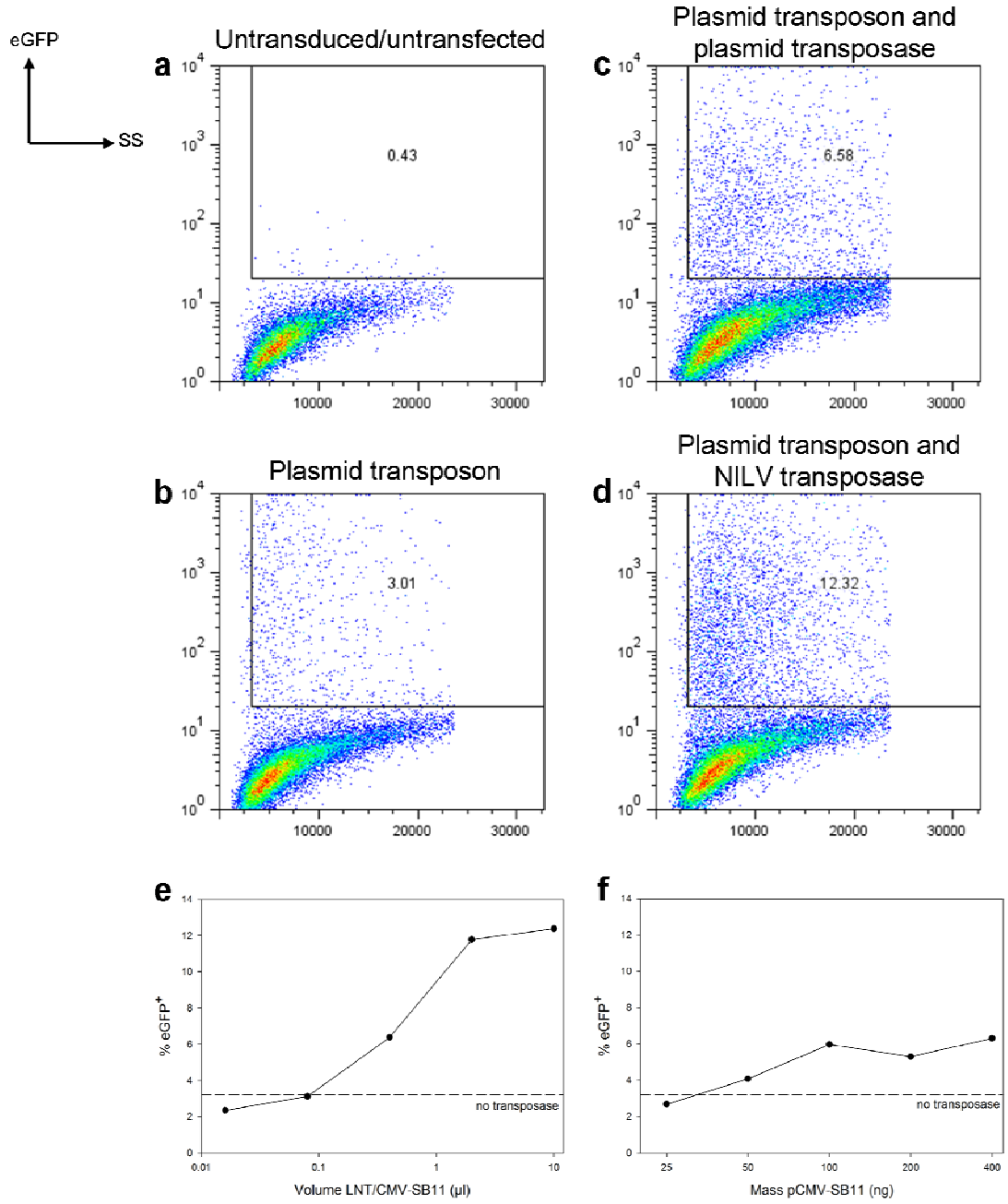

Figure 4.8 Transposition from pLNT/tn[SFFV-eGFP] FWD.

10^5 293T cells were transfected with 400ng pLNT/tn[SFFV-eGFP] FWD and either transduced with non-integrating LNT/CMV-SB11 lentiviral vector or transfected with pCMV-SB11. eGFP expression was determined by flow cytometry 14 days post-transfection/transduction. The proportion of cells expressing eGFP is given as a percentage within the gated area. **(a)** untransfected/untransduced; **(b)** 400ng pLNT/tn[SFFV-eGFP] FWD; **(c)** 400ng pLNT/tn[SFFV-eGFP] FWD and 400ng pCMV-SB11 **(d)** 400ng pLNT/tn[SFFV-eGFP] FWD and 10µl LNT/CMV-SB11. e and f show the effect of titrating the volume of LNT/CMV-SB11 and mass of pCMV-SB11 respectively. NILV, non-integrating lentiviral vector.

4.3.8. Transposition from a Lentiviral Vector

The integration assay was next used to determine the ability of non-integrated lentiviral DNA to act as a transposon donor. When 293T cells were transduced with non-integrating LNT/tn[SFFV-eGFP] FWD at a functional multiplicity of infection (MOI) of 1 and transduced or transfected with serial dilutions of transposase expression lentiviral vector or plasmid respectively, eGFP persistence at day 14 was almost undetectable under all conditions tested (data not shown).

In order to test the possibility that increasing the transposon vector MOI would increase the level of transposition, 293T cells were transduced with nonintegrating LNT/tn[SFFV-eGFP] FWD at a functional MOI of 20 and co-transfected with a transposase expression plasmid (Figure 4.9). The initial level of gene transfer was extremely high, with over 96% of cells expressing eGFP 2 days post-transduction. After 14 days, the proportion of $eGFP^+$ cells was approximately 2-fold higher in cells co-transfected with transposase expression plasmid than in those which were not co-transfected, suggesting that the non-integrated lentiviral DNA may have been acting as a substrate for transposition, albeit inefficiently. However, the low increase in eGFP persistence in this experiment combined with the lack of repetitions means that the statistical power of this result is not strong.

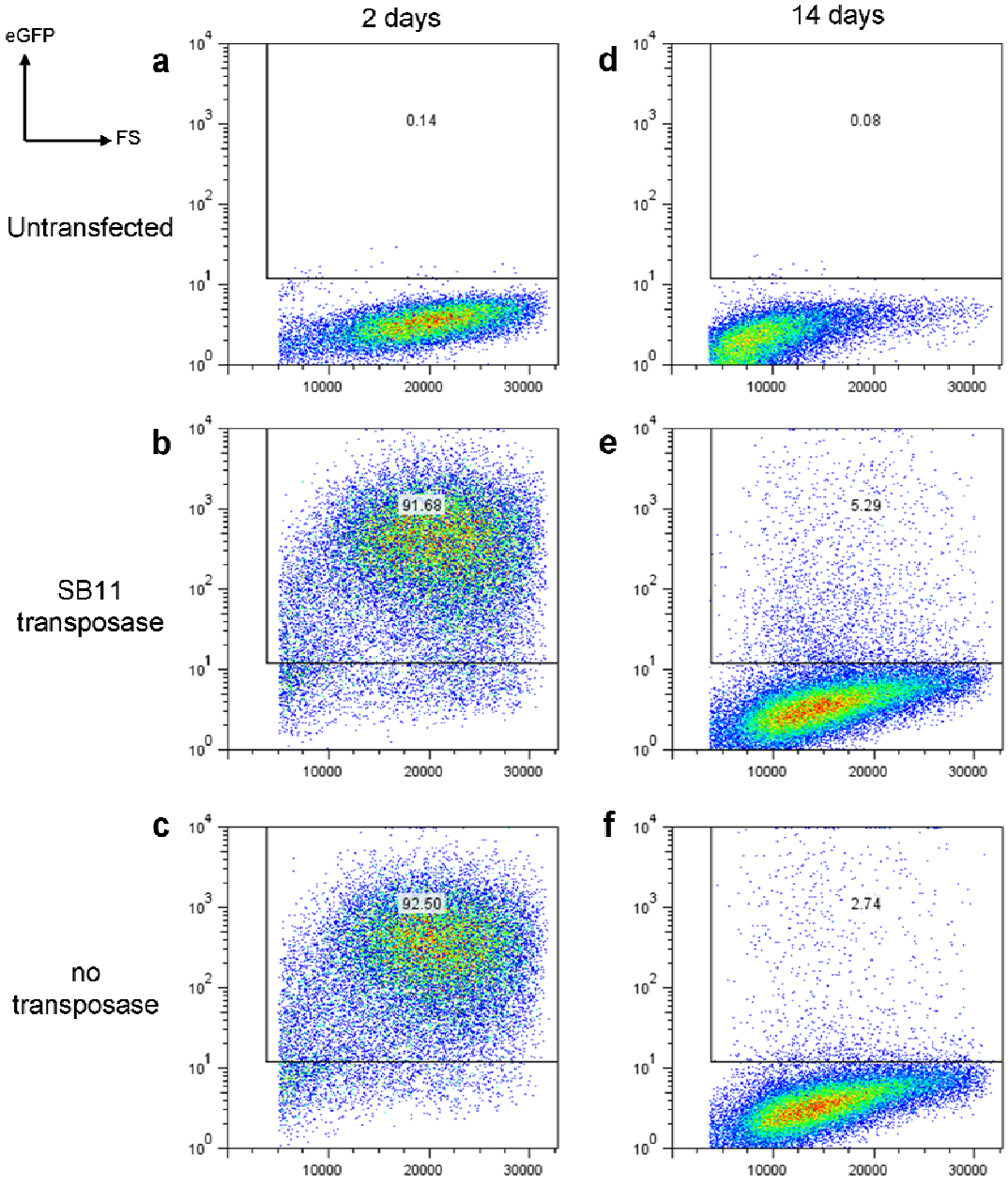

Figure 4.9 Transposition from an eGFP transposon lentiviral vector.

10^5 293T cells were either untransduced/untransfected **(a, d)**, transduced at MOI 20 with LNT/tn[SFFV-eGFP] FWD and co-transfected with 400ng pCMV-SB11 **(b, e)**, or not co-transfected **(c, f)**. eGFP expression was determined by flow cytometry 2 days **(a-c)** and 14 days **(d-f)** post-transfection/transduction. The proportion of cells expressing eGFP is given as a percentage within the gated area.

4.3.9. Analysis of Hybrid Vector Integration by Ligation-Mediated PCR

The most direct evidence for transposition is provided by cloning and sequencing transposon-chromosome junctions from target cells. This was performed on the eGFP^{+}-sorted DNA samples by ligation-mediated PCR (Figure 4.10). The genomic DNA was digested with NlaIII, a restriction enzyme which has a 4bp recognition sequence and so cuts genomic DNA frequently. The digested DNA was ligated to synthetic linker DNA and subjected to nested PCR using primer pairs which bind in the linker DNA and the transposon right inverted repeat. This allows the amplification of transposon-chromosome junctions for subsequent plasmid cloning and sequencing.

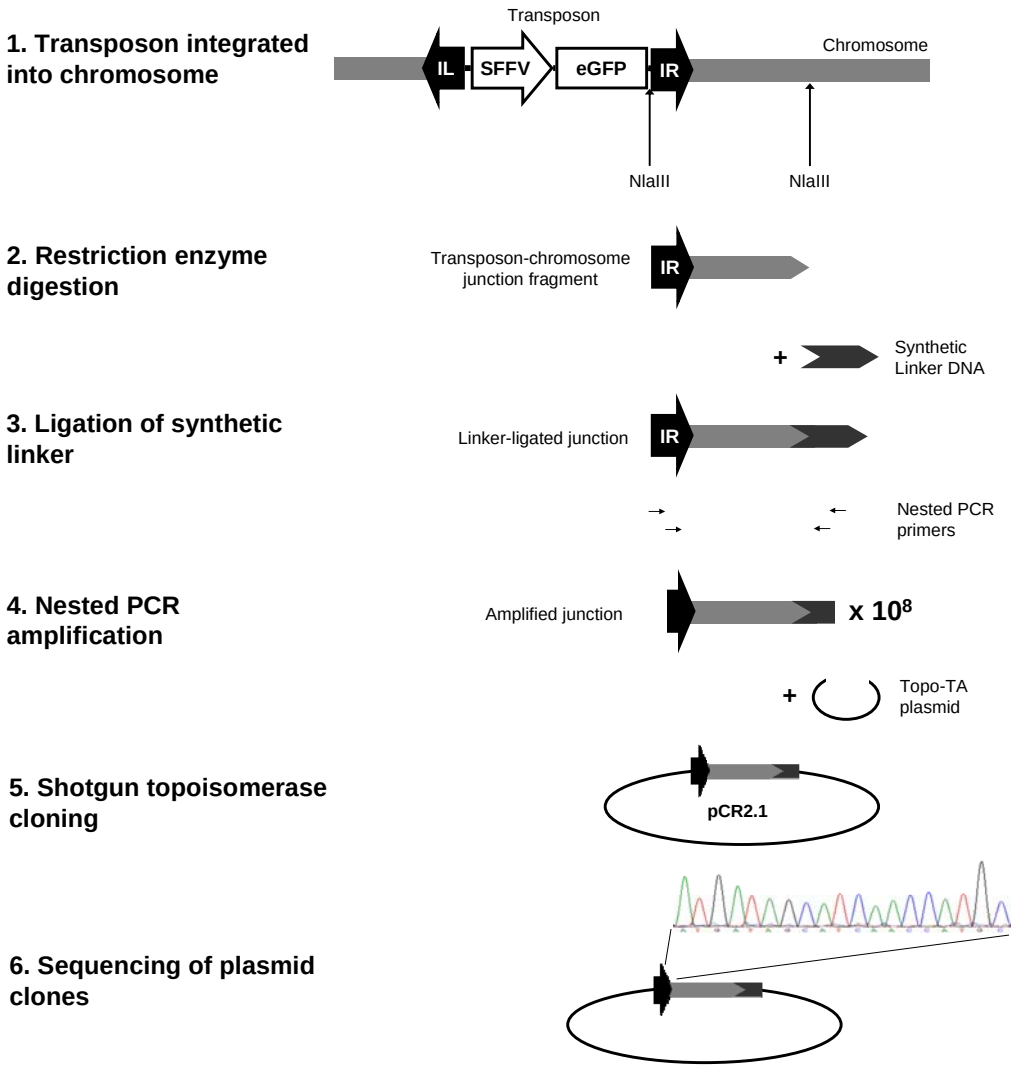

Figure 4.10 Ligation-mediated PCR schematic overview.

(1) DNA containing transposon-chromosome junctions is **(2)** digested with restriction enzymes to produce smaller fragments with known overhangs. **(3)** These are ligated to an excess of synthetic linker DNA with matching overhangs, resulting in transposon-chromosome junctions flanked by a known sequence at both ends. **(4)** These junctions are amplified by nested PCR and **(5)** shotgun-cloned using topoisomerase-linked plasmids. **(6)** The plasmids are recovered and made clonal by bacterial transformation and plasmid DNA is extracted and sequenced. This sequence can be compared to genome databases in order to identify the location of transposon integration.

Transposon Donor	Junction Sequence (see legend for colour code)	Integration Type	Integration Location
Plasmid 1	TAAACTTCCGACTTCAACTGTAGCTTGATTTCTACAAAATTCTTAA GTTCTGAATGAGACTTTGGTTTTAACTCACCCAAAAATTCAGCATG TGAGTTTCG	Transposition	Chromosome 8, + strand, 125,589,969-125,590,049
Plasmid 2	TAAACTTCCGACTTCAACTGTATCTTCCCATAAAAACTAGACAGAA GCATTCTCAGAGACTTATTTGTGATGTGTGTCCTCAACTAACAGAG TTGAACCTTTCTTTTGATACAGCAGTTTGGAAACACTCTTTTTGTA	Transposition	Repeat sequence
Plasmid 3	TAAACTTCCGACTTCAACTGTAGATTCTAACTGGAAGAGAGCCTCA GAAAGGACCTCAGAAAGGGGATGGGAGGGGTATATATGTATATCTA TCAAATATCA	Transposition	Chromosome 11, - strand, 99,947,423-99,947,502
Plasmid 4	TAAACTTCCGACTTCAACTGTATTAGTTGAGATCATGGGAATTTAA GAAATAGGAACTGGGCAGGTGCGGTGGCTCACGCCTCTAATCTCAG CATTTTGGGAGGCCGAGGCGGGTGAATCACCTGAGCCCAGCAGTTC	Transposition	Chromosome 3, - strand, 152,607,497-152,607,614
Plasmid 5	TAAACTTCCGACTTCAACTGTACATTGTCG	(Transposition)	Unknown – sequence too short to identify
NILV 1	TAAACTTCCGACTTCAACTGTATAGGGATCCGAATTCG	Background	Not mobilised
NILV 2	TAAACTTCCGACTTCAACTGTATAGGGATCCGAATTCGA	Background	Not mobilised
NILV 3	TAAACTTCCGACTTCAACTGTATAGGGATCCGAATTCGA	Background	Not mobilised
NILV 4	GGTACCAAGCTTGATGCATAGCTTGAGTATTCTATAGTGTCACCAA AATAGCTTGGCGTAATCATGGTCATAACTGTTTCCTGTGTGAAATT GTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAA	DNA not identifiable	Unknown
NILV 5	AAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACCTTA ATTGCCTTGCCCTCA	DNA not identifiable	Unknown

Table 4.1 Integration sites recovered from eGFP$^+$-sorted DNA.

10^5 293T cells were transfected with 400ng pCMV-SB11 and transduced with LNT/tn[SFFV-eGFP] FWD at MOI 20 (NILV sites) or co-transfected with 400ng pT2/SFFV-eGFP (Plasmids sites). Cells were passaged for 14 days and the eGFP$^+$ population for each condition was isolated by fluorescence-activated cell sorting. These populations were expanded and genomic DNA was extracted. Integration sites were recovered from this DNA by ligation-mediated PCR and sequenced. The DNA flanking the transposon inverted repeat was identified by Blat searching the *hg18* March 2006 database hosted by the UCSC Genome Server. Red, transposon inverted repeat; black underlined, TA junction; blue, chromosomal DNA; green, lentiviral backbone; grey, unknown.

Plasmids containing LM-PCR fragments were screened by PCR and five from each condition were sent for sequencing of the junction. The sequencing results are shown in Table 4.1. All five of the junctions cloned from $eGFP^+$ plasmid transposon DNA showed the three essential features of a successful transposition event. Firstly, the full inverted repeat sequence was present. Secondly, the inverted repeat was immediately upstream of a TA-dinucleotide. Thirdly, the TA-dinucleotide was immediately upstream of chromosomal DNA (as identified by a Blat search on the UCSC Genome Server), and a TA-dinucleotide was present at the correct position in the human genome sequence database. By contrast, three of the $eGFP^+$ lentivirus-transposon DNA junctions showed that the transposon had not moved from its initial position in the lentiviral backbone. The other two cloned inserts contained DNA which could not be identified by Blat search or comparison with the lentiviral backbone.

4.4. Summary

- An eGFP-based integration assay was established using plasmid transfection. Transposon integration into chromosomes was confirmed by ligation-mediated PCR.
- High titre non-integrating lentiviral vectors carrying the eGFP transposon could be prepared using constructs in which the transposon was in the same direction as the lentiviral backbone.
- Transient expression of transposase from non-integrating lentiviral vectors was demonstrated by Western blot. These vectors were able to drive transposition from plasmids in the eGFP integration assay.
- When cells were transduced with a lentivirus-transposon vector at high MOI and transfected with a transposase expression plasmid, there was a slightly increased persistence of eGFP expression relative to background integration. However, the low level of signal and weak statistical power of the experiment did not clearly support the hypothesis that transposon mobilisation from a lentiviral backbone was taking place.
- Transposon mobilisation from a lentiviral backbone was not detected by ligation-mediated PCR.

In this chapter it was shown that lentiviral vector delivery of Sleeping Beauty transposons and transposase to target cells is feasible. In cells transduced at a high MOI with a non-integrating lentiviral vector carrying a transposon, the rate of integration increased 2-fold in the presence of transposase. However, the absolute efficiency of this reaction was low, resulting in 5.3% of cells becoming stably $eGFP^+$ with transposase compared to 2.7% without. The lack of statistical power in this experiment meant that the strong conclusion of transposition from a lentiviral backbone could not be drawn. In addition, integration sites resulting from Sleeping Beauty transposition could not be recovered from these cells by ligation-mediated PCR.

A possible solution to this problem would be to test a wider range of conditions to see if there are more optimal circumstances for transposition out of a lentiviral backbone. For example, the transposon and transposase vector dose can be titrated against one

another to identify the optimal ratio. In the literature, there is significant variation between authors in terms of transposase expression vector designs, transfection conditions, and methods of reporting transposition efficiency. Some authors report optimisation methods such as a 1:1 or 20:1 ratio of transposon to transposase plasmid in their transfection mixes, though presumably the final number of plasmids per target cell and the ultimate level of transposase expression within each cell can vary greatly depending on the method of transfection, the cell type, and the transposase expression plasmid being used.

Another solution would be to develop a more sensitive assay for chromosomal integration. Persistence of eGFP expression on day 14 is not ideal as fluorescent cells may contain only episomal DNA or eGFP protein (though the latter is unlikely to contribute significantly as its half life in mammalian cells is about 26 hours (493)). Perhaps more importantly, the efficiency of Sleeping Beauty transposition is such that even under the best conditions for transposition there is a large proportion of cells (90-99%) in which no integration event takes place. The presence of these cells in the population makes it more difficult to analyse integration by other means (such as LM-PCR) as it increases the sensitivity requirement of these other assays.

One way to exclude cells not containing an integration event from the LM-PCR reaction would be to use a reporter gene which confers resistance to an antibiotic such as neomycin. Adherent cells are transduced or transfected as appropriate and seeded at low density under selective conditions. Cells in which integration has taken place give rise to resistant colonies while the rest are lost. While antibiotic selection cannot distinguish between background integration and transposition, it is a much more efficient method of generating genomic DNA containing integration events than cell sorting and so makes it practical to screen a larger number of samples for transposition events.

Given the data and theoretical considerations presented in this chapter, it was decided to replace the eGFP reporter in the lentivirus-transposon vector with a neomycin phosphotransferase reporter which confers resistance to the antibiotic neomycin. In addition, it was decided to screen a large number of conditions in which the amounts of transposon and transposase were varied in an attempt to identify more optimal conditions for transposition out of a lentiviral backbone, as well as to perform

multiple repetitions of each condition in order to increase the statistical power of the resulting data.

5. Lentivirus-Transposon Hybrid Vectors Conferring Antibiotic Resistance

5.1. Aims

- To produce a high titre lentivirus-transposon vector expressing the antibiotic resistance gene neomycin phosphotransferase
- To optimise the conditions for lentivirus-transposon integration
- To confirm transposition out of a lentiviral backbone by cloning and sequencing transposon-chromosome junctions
- To determine the integration profile of lentivirus-transposon vector integration

5.2. Introduction

The primary chromosomal integration assay used in the previous chapter was based on persistence of eGFP expression after passaging cells for two weeks following gene transfer. The assay hinges on the assumption that two weeks of cell division will dilute episomal DNA to the point where it no longer contributes significantly to reporter gene expression within the population.

In the work described in this chapter, the eGFP reporter gene within the transposon was replaced with a neomycin phosphotransferase expression cassette, which confers resistance to the antibiotic G418. In the revised assay, cells were transduced or transfected as appropriate, cultured under non-selective conditions for 3 days to permit reporter gene expression, and then seeded at low density into fresh plates in triplicate and cultured with G418 selection for 2 weeks. Cells containing a stably integrated neomycin phosphotransferase expression cassette are able to divide under selective conditions and give rise to resistant colonies, while cells lacking the transgene or containing only episomal DNA do not produce colonies.

This G418-resistant colony assay is based on the same assumption as the eGFP persistence assay, namely that continued gene expression after 2 weeks of cell division must be due to integrated vector DNA. However, it has two important advantages over the previous system. Firstly, it should be better able to distinguish

between persistent integrated and episomal DNA: although both types are able to confer G418-resistance on individual target cells, those carrying only episomal DNA cannot give rise to colonies. Secondly, selection against cells which do not contain integrated transgene results in a cell population which is greatly enriched for integration events. This enrichment may reduce the level of sensitivity required for recovery of integration sites by LM-PCR.

This chapter describes the production of a lentivirus-transposon vector expressing neomycin-phosphotransferase, the optimisation of its use, the recovery of hybrid vector integration sites, and bioinformatic analysis of these sites to determine the integration profile relative to the parental lentiviral and plasmid-based systems.

5.3. Production and Testing of Lentivirus-Transposon Vectors Expressing Neomycin Phosphotransferase

5.3.1. Plasmid subcloning of pLNT/tn[SV40-Neo]

A Sleeping Beauty transposon containing a simian virus 40 promoter and neomycin phosphotransferase coding sequence was obtained from plasmid pT2/SV40-Neo (494) (a kind gift from Scott McIvor, University of Minnesota) by BamHI digestion. The fragment was ligated into the lentivirus-transposon backbone described in Section 4.3.1 at the BamHI site so that the neomycin phosphotransferase transposon replaced the eGFP transposon. Colonies were screened for transposons inserted in the forward orientation as shown in Figure 5.1.

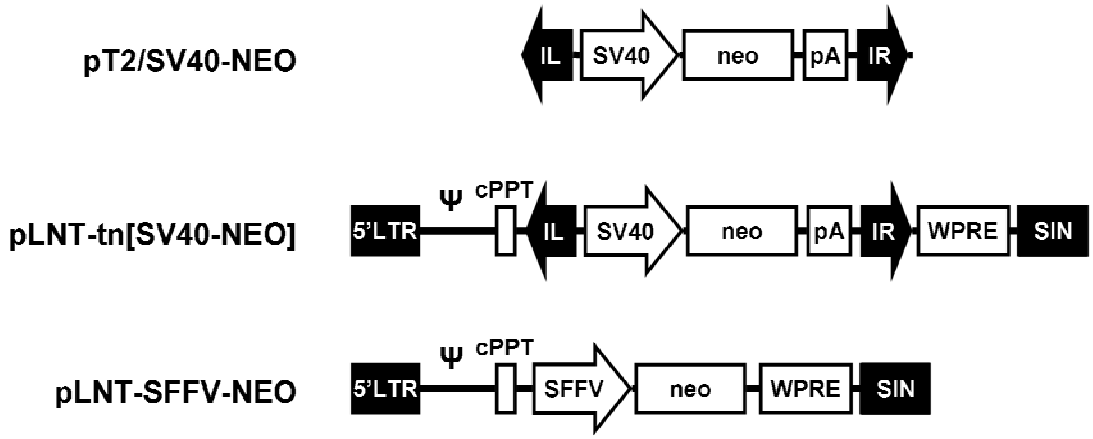

Figure 5.1 Vector map of neomycin transposon and lentiviral plasmids.

SV40, simian virus 40 promoter; neo, neomycin phosphotransferase. pA, polyadenylation signal from simian virus 40; 5'LTR, HIV-1 5' long terminal repeat; Ψ, HIV-1 RNA packaging signal; SIN, self-inactivating (U3-deleted) HIV-1 long terminal repeat; cPPT, central polypurine tract; WPRE, woodchuck posttranscriptional regulatory element; IL, Sleeping Beauty transposon left inverted repeat; IR, Sleeping Beauty transposon right inverted repeat; SFFV, spleen focus forming virus LTR promoter.

5.3.2. Production and Titration of Neomycin Resistant Hybrid Vectors

In order to test expression from this vector, integrating lentiviral vector was produced in parallel using pLNT/tn[SV40-Neo] and a standard pLNT/SFFV-Neo construct (a kind gift from Luis Apolonia, University College London). HeLa cells were used to assess G418 resistance because 293T cells were originally engineered using a neomycin phosphotransferase construct (495) and initial experiments with HT1080 cells resulted in large, diffuse colonies. HeLa cells were transduced with serial dilutions of the two vectors, passaged without selection for 3 days, trypsinised and seeded at low density into 6 well plates and incubated in medium containing G418 for 2 weeks to produce cell colonies. The plates were fixed with paraformaldehyde, stained with crystal violet, and imaged using a conventional office flatbed scanner (Figure 5.2). By counting the resistant colonies, the integrating titre of the LNT/tn[SV40-Neo] construct was estimated to be 2.88 x 10^8 TU/ml versus 9.91 x 10^8 TU/ml for LNT/SFFV-Neo (Figure 5.3a).

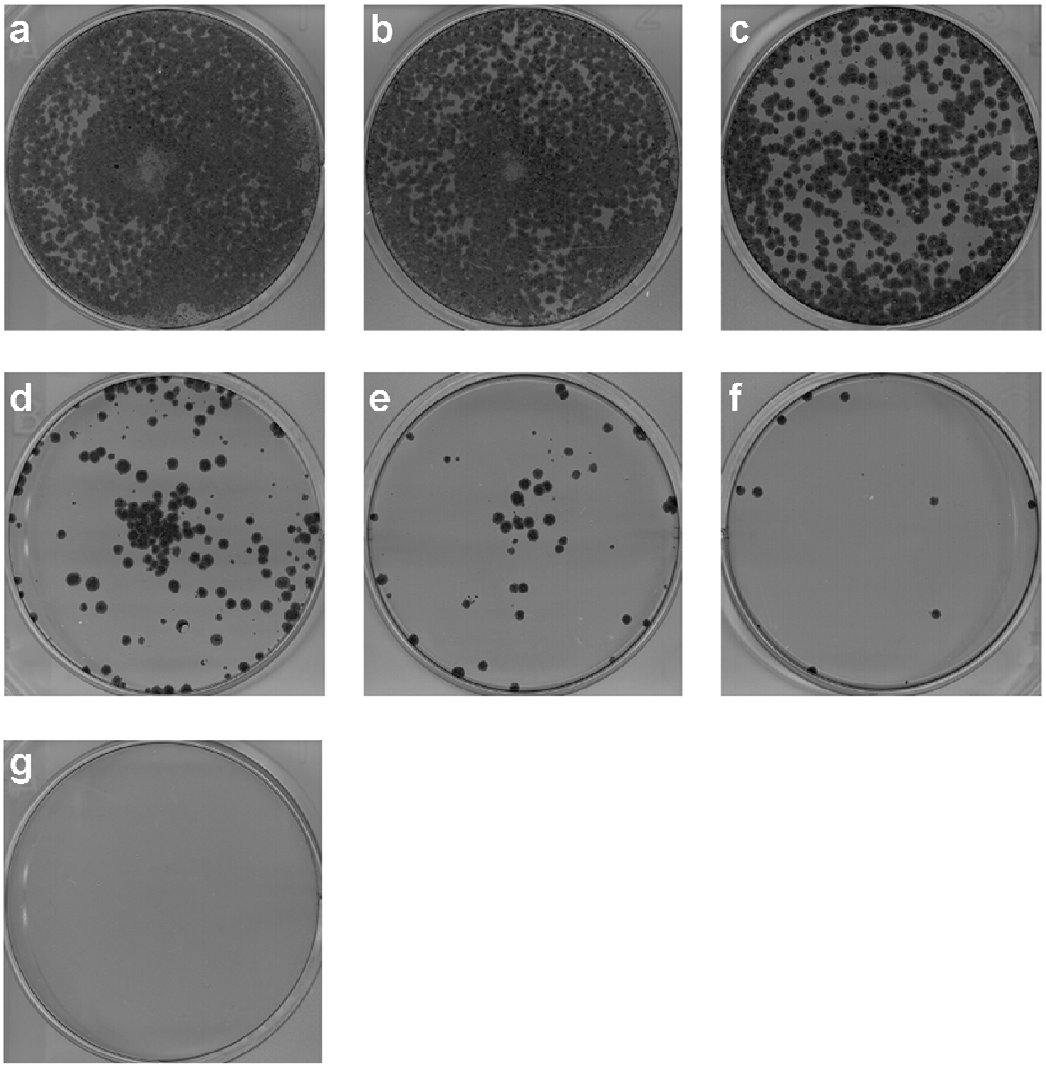

Figure 5.2 G418-resistant HeLa cell colonies following transduction with integrating lentivirus-transposon vector.

Lentiviral vector was produced using pLNT/tn[SV40-Neo] and an integrase-proficient packaging plasmid. 10^5 HeLa cells were transduced with serial dilutions of vector, incubated for 3 days without selection, trypsinised and 1/10th of the cell suspension was seeded into a fresh 6-well plate. Cells were maintained in medium containing 1mg/ml G418 for two weeks before fixing and staining. **(a)** 1µl LNT/tn[SV40-Neo]; **(b)** 0.2 µl; **(c)** 0.04 µl; **(d)** 0.008 µl; **(e)** 0.0016 µl; **(f)** 0.00032 µl; **(g)** untransduced.

Next, non-integrating LNT/tn[SV40-Neo] and LNT/CMV-SB11 were produced in parallel. HeLa cells were transduced with serial dilutions of both vectors. 3 days after transduction total DNA was extracted for vector titration by quantitative PCR using primers and probes in the vector WPRE and the host β-actin. The estimated titres of LNT/tn[SV40-Neo] and LNT/CMV-SB11 were 4.02 x 10^8 TU/ml and 4.68 x 10^8 respectively (Figure 5.3b). HeLa cells are a genetically stable but aneuploid cell line (496), so the β-actin copy number within a given sample is expected to be proportional to but not double the value of the cell number as would be expected for normal diploid cells.

The physical titre of the two vectors was measured by HIV-1 p24 ELISA. The estimated physical titres of LNT/tn[SV40-Neo] and LNT/CMV-SB11 were 2.85 x 10^8 pg p24/ml and 2.57 x 10^8 pg p24/ml respectively (Figure 5.3c).

Together these titration data suggest that the hybrid neomycin phosphotransferase vector could be produced at similar titres to standard ultracentrifuge-concentrated lentiviral vectors.

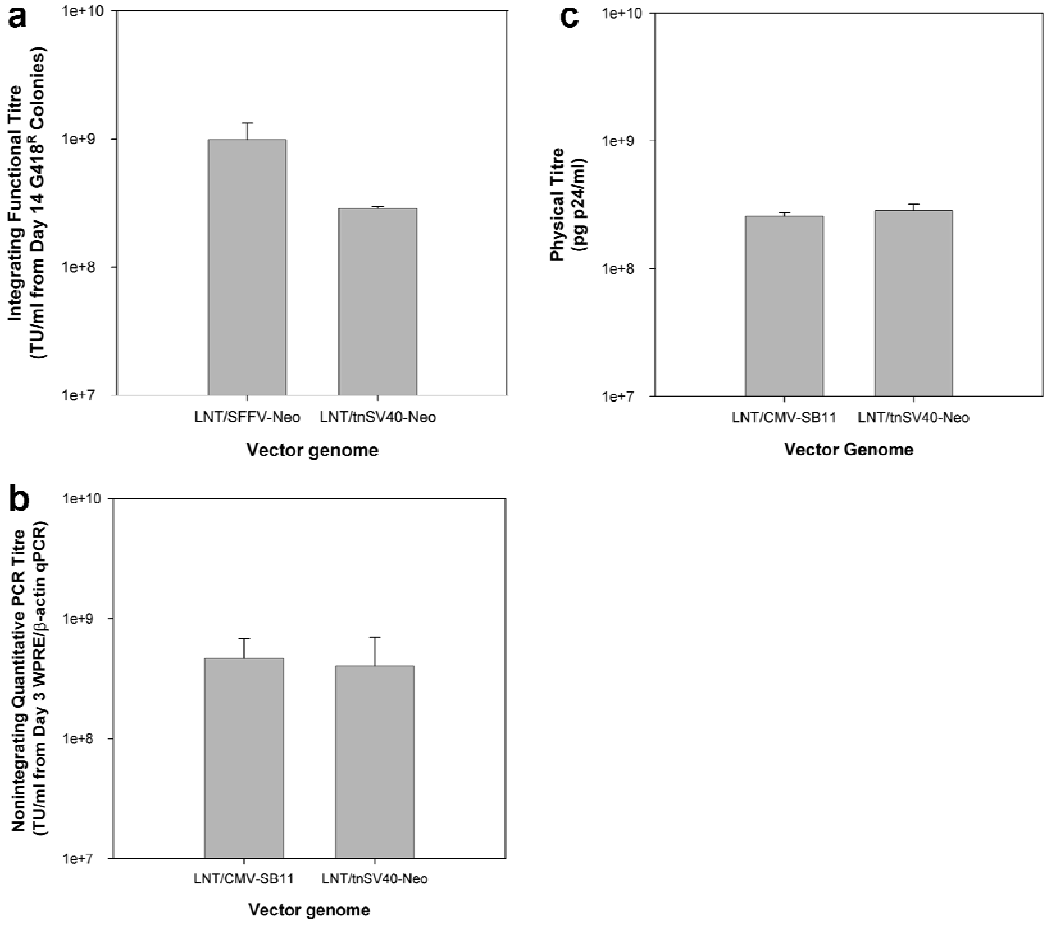

Figure 5.3 Titration of neomycin resistant lentivirus-transposon vectors.

(a) Integrating functional titre. 10^5 HeLa cells were transduced with serial dilutions of integrating LNT/SFFV-Neo or integrating LNT/tn[SV40-Neo]. 3 days post-transduction the cells were trypsinised and $1/10^{th}$ of the cell suspension seeded into a fresh 6-well plate. The cells were incubated in medium containing 1mg/ml G418. 2 weeks post-transduction colonies were fixed, stained, and counted. **(b)** Non-integrating quantitative PCR titre. 10^5 HeLa cells were transduced with serial dilutions of nonintegrating LNT/CMV-SB11 or LNT/tn[SV40-Neo]. Total DNA was extracted 3 days post-transduction and the copy number of the viral WPRE or the host cell β-actin determined by quantitative PCR. Titres are expressed as the ratio of the WPRE copy number to the β-actin copy number for each sample. **(c)** Non-integrating physical titre. The concentration of the HIV-1 p24 antigen in concentrated preparations of non-integrating LNT/CMV-SB11 or LNT/tn[SV40-Neo] was assessed by ELISA.

5.3.3. Automated Mammalian Cell Colony Counting

Initial experiments using G418-resistant HT1080 colonies showed that mammalian cell colonies can be heterogeneous in shape and size, making accurate and unbiased manual counting difficult. The process is also labour-intensive and counts may vary between operators or even the same operator at different times. This places limits on the number of conditions and replicates that can be tested at any one time.

In order to address this problem, a protocol for automated colony counting was established using the CellProfiler software platform in Matlab ((497) and Figure 5.4). Cell colonies were fixed with paraformaldehyde, stained with crystal violet, and imaged using a conventional office flatbed scanner. In order to maintain consistency between scanned images, a card template was produced which sits in the scanner bed, thus ensuring that well locations were consistent between plates. Image files were loaded by the CellProfiler software and well locations were manually described via point-and-click. Each well image was automatically cropped and inverted, and the background intensity was manually defined. Colonies were identified by size, shape, intensity, and separation using user-defined values, and the results were displayed graphically for verification by the user (Figure 5.4e). The combination of standardised image processing and visual verification enabled high throughput counting of mammalian colonies with minimal user-introduced bias.

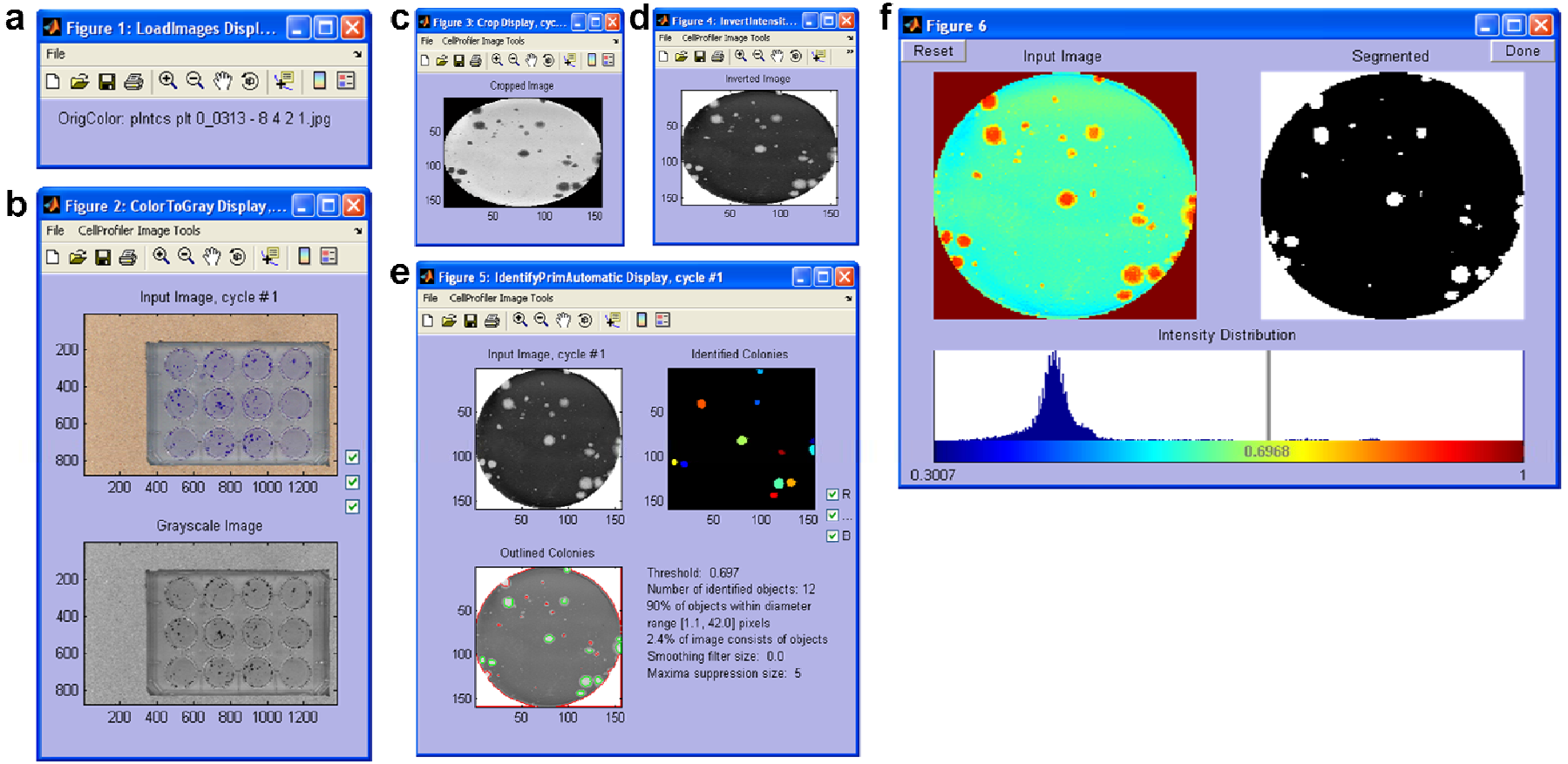

Figure 5.4 Cell colony counting using CellProfiler software.

Before counting, colonies are fixed with paraformaldehyde and stained with crystal violet, then dried and imaged using a conventional office scanner. **(a)** loading image files; **(b)** converting images to grayscale; **(c)** cropping the image to single wells; **(d)** inverting the well image to show colonies in white; **(e)** automated identification and counting of colonies based on shape, size, separation, and staining intensity; **(f)** setting the background intensity manually.

5.3.4. Relating the Level of Transposition to the Initial Transposon Number

For a hybrid vector to combine the gene transfer mechanism of one vector with the integration mechanism of another, the DNA output of the gene transfer step must be compatible with the DNA input of the integration step. Many factors may influence the ability of a gene transfer vector to act as a substrate for Sleeping Beauty transposition, such as the physical state of the DNA (e.g. whether it is supercoiled, circular, or linear (498)), its subcellular location (e.g. cytoplasmic or nuclear), or copy number.

In order to investigate the relationship between episomal transposon copy number and the level of transposition, HeLa cells were transfected with the transposase expression plasmid pCMV-SB11 or mutant transposase plasmid pCMV-SBΔDDE and co-transfected or co-transduced with a serial dilution of transposon plasmid pLNT/tn[SV40-Neo] or non-integrating lentiviral transposon vector LNT/tn[SV40-Neo]. After three days, cells were trypsinised and samples taken to determine the vector copy number per cell by quantitative PCR. A fraction of each cell suspension was seeded at low density into fresh 6-well plates and cultured with G418 for 14 days. Colonies were fixed, stained, and counted using CellProfiler.

The results are shown in Figure 5.5 below. When cells were transfected with small masses of plasmid (transfection plasmid masses below 20ng per 10^5 cells) so that the measured transposon copy number was less than 30, there was little observable difference in the rate of integration in the presence or absence of active transposase. However, at higher copy numbers in the range of 500 to 10,000 plasmids per cell, expression of active transposase increased the level of integration 8-fold relative to cells containing inactive transposase.

When transposons were delivered by non-integrating lentiviral vector, there was no observable change in the level of integration in the presence of transposase, even at the highest transposon copy number of 20. It is possible that the much lower copy numbers of NILV compared to plasmids meant that there was insufficient transposon substrate within cells for a detectable level of transposition to take place for this vector. In addition, the rate of NILV background integration in the absence of

transposase activity was much greater than that observed for plasmid. It has been reported that linear episomal DNA is integrated by host DNA repair pathways into chromosomes more efficiently than circular DNA (499).

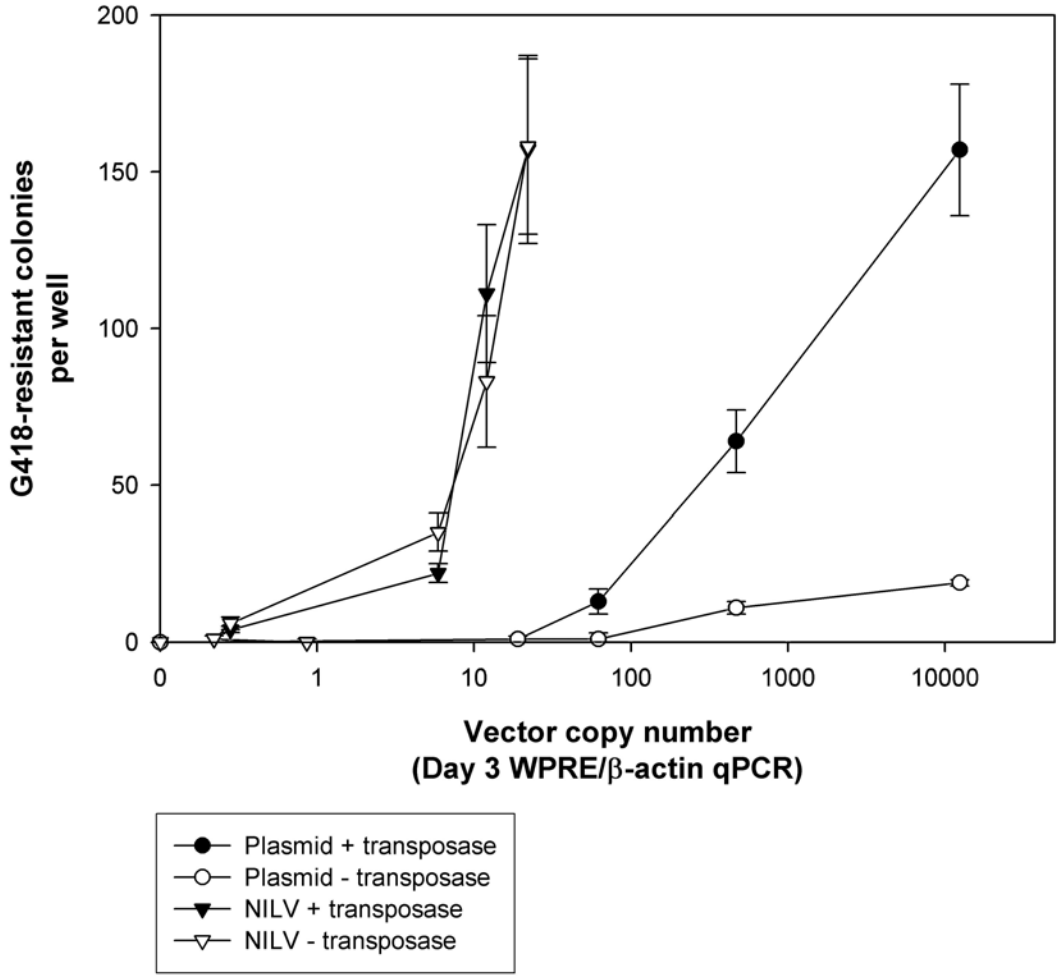

Figure 5.5 The effect of transposon copy number on the rate of transposition.

10^5 HeLa cells were transfected in triplicate with 400ng active transposase plasmid pCMV-SB11 (+ transposase) or 400ng inactive mutant transposase pCMV-SBΔDDE (- transposase) and either co-transfected with serial dilutions of pLNT/tn[SV40-Neo] or co-transduced with serial dilutions of non-integrating LNT/tn[SV40-Neo]. On day 3 post-transduction, cells were trypsinised and $1/10^{th}$ of the cell suspension was seeded into a fresh 6-well plate. Cells were incubated for 2 weeks in medium containing 1mg/ml G418 and colonies were fixed and stained for counting. The copy number of plasmid or lentiviral genome was determined on day 3 by WPRE/β-actin quantitative PCR using the SB11-transfected samples. Vector copy numbers are expressed as the ratio of the WPRE copy number to the β-actin copy number for each sample.

5.3.5. Optimisation of Transposase Delivery

Optimisation of transposase expression has been reported to enhance the rate of transposition by ensuring the presence of sufficient transposase protein for the transposition reaction while avoiding 'overproduction inhibition', a reported decline in transposition efficiency when transposase is overexpressed (500). In order to determine the sensitivity of the transposition reaction to the level of transposase expression, HeLa cells were transfected with pLNT/tn[SV40-Neo] and transduced with serial dilutions of nonintegrating LNT/CMV-SB11 which had been previously titred by quantitative PCR. A G418 colony assay was performed as previously described. The results are shown in Figure 5.6.

The rate of transposition generally increased with transposase NILV copy number, reaching a maximum level of 300 colonies per well (2.5% gene marking above background) at a transposase vector copy number of 29 (corresponding to 4µl of lentiviral vector preparation per 10^5 cells).

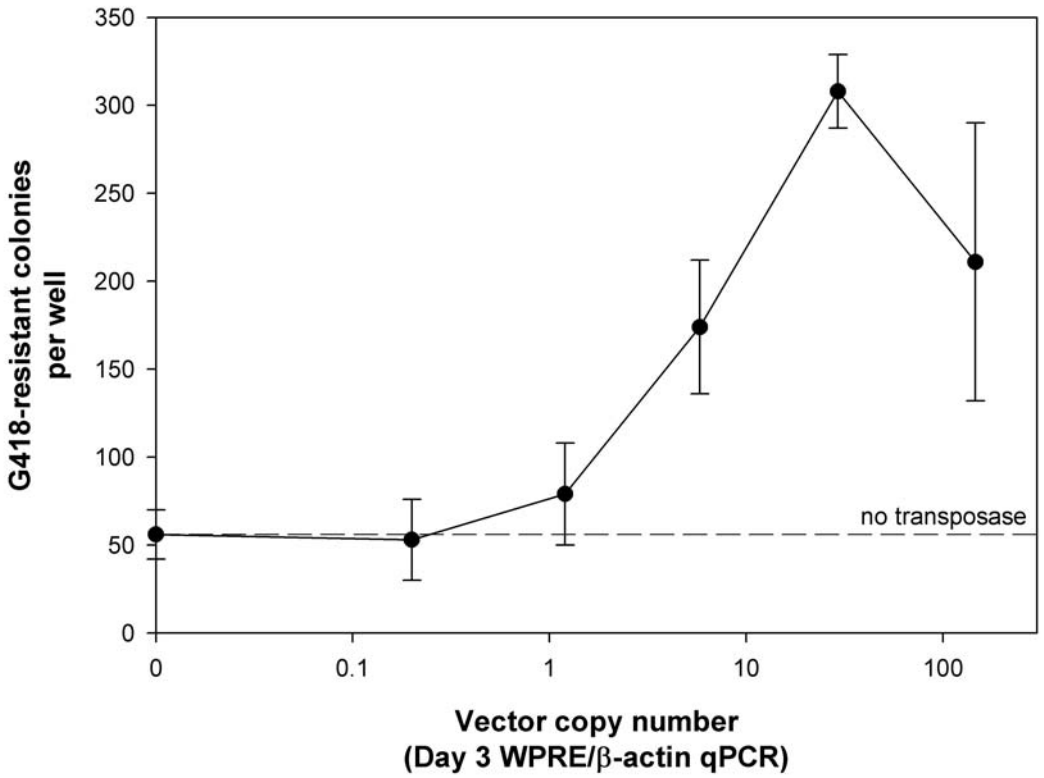

Figure 5.6 The effect of transposase optimisation on transposition efficiency.

10^5 HeLa cells were transfected in triplicate with 400ng pLNT/tn[SV40-Neo] and transduced with serial dilutions of nonintegrating LNT/CMV-SB11. 3 days post-transduction the cells were trypsinised and $1/10^{th}$ of the cell suspension seeded into a fresh 6-well plate. Cells were maintained in medium containing 1mg/ml G418 for 2 weeks before fixing and staining for colony counting. The transposase MOI was determined by WPRE/β-actin quantitative PCR using non-transfected HeLa cells. Vector copy numbers are expressed as the ratio of the WPRE copy number to the β-actin copy number for each sample.

As both the transposon and transposase vector copy numbers were found to affect the rate of transposition, a double titration of both the transposon and transposase components was next performed on the basis that this approach was the most thorough way to identify the optimal conditions for transposition out of a lentiviral backbone.

HeLa cells were transfected or transduced with serial dilutions of both the transposon and transposase components and integration was measured using the G418-resistant colony assay. Three conditions were tested: plasmid transposon, plasmid transposase; plasmid transposon, NILV transposase; NILV transposon, NILV transposase. To ensure a high level of gene transfer, the maximum plasmid mass used was 8 μg per 10^5 cells and the maximum lentiviral vector volume was 80μl containing 2.4μg of p24 antigen. It is worth noting that this latter volume is approaching the maximum that can be achieved using concentrated lentiviral vector in a single transduction because transductions were performed in a total volume of 500μl per well.

The results are shown in Figure 5.7. When both Sleeping Beauty components were transfected on plasmids (Figure 5.7a,b), the highest rate of gene marking observed was 1.4% ± 0.4% at 4μg pLNT/tn[SV40-Neo] and 1μg pLNT/CMV-SB11 per 10^5 cells. The background (no transposase) rate was 0.4% ± 0.2% for this condition ($p<0.02$). When transposase NILV was provided with transposon plasmid (Figure 5.7c,d), the highest rate of gene marking was 3.1% ± 0.4% (background 0.5% ± 0.2%, $p<0.0005$) at 2μg pLNT/tn[SV40-Neo] and 0.52μg p24 LNT/CMV-SB11 per 10^5 cells. When both components were delivered by NILV transduction (Figure 5.7e,f), the highest rate of gene marking was 2.6% ± 0.2% (background 1.4% ± 0.6%, $p<0.03$ by 2-tailed t-test) obtained at 2.4μg p24 LNT/tn[SV40-Neo] and 0.13μg p24 LNT/CMV-SB11 per 10^5 HeLa cells. The rate of gene marking detected at the other transposase NILV concentrations tested was not significantly different from the rate of background integration at the 5% level, although the 0.03μg and 0.5μg p24 LNT/CMV-SB11 conditions were significant at the 10% level.

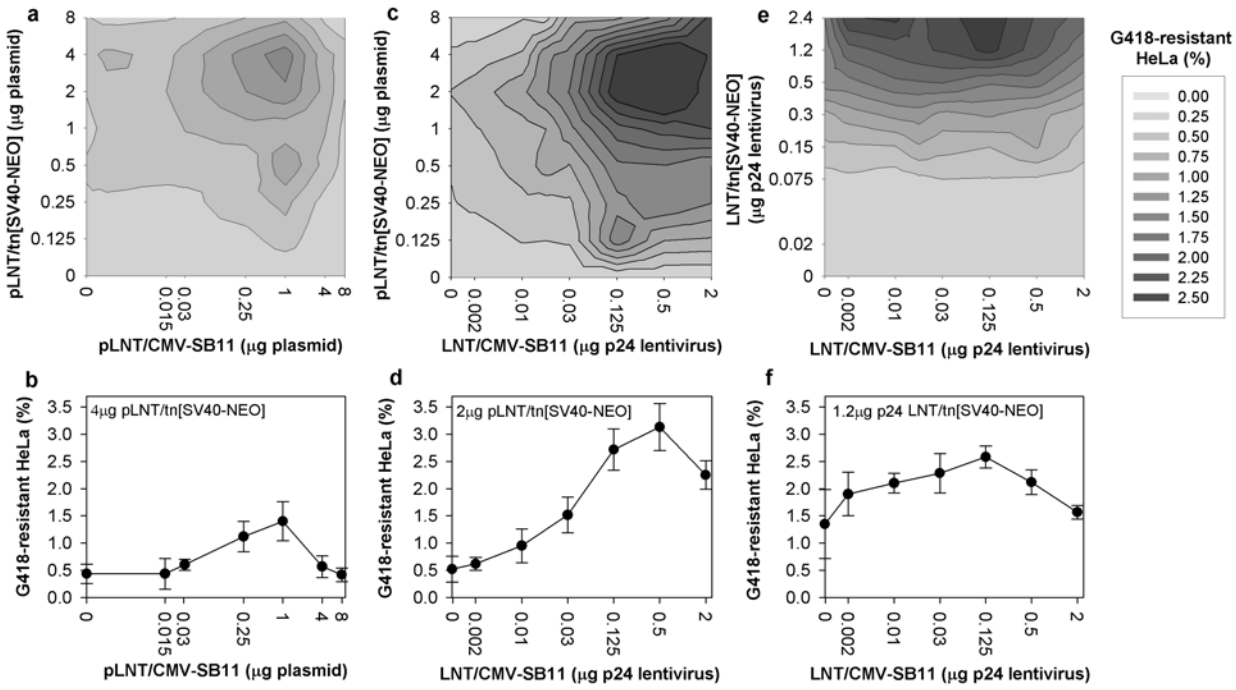

Figure 5.7 Double titration optimisation of transposon and transposase delivery.

10^5 HeLa cells were transduced or transfected with Sleeping Beauty components in a double titration integration assay. The transposon amounts tested are given by the y-axis labels in a, c, and e, while the transposase amounts are given by the x-axis labels. All transposon-transposase combinations were tested, and each combination was tested in triplicate. **(a)** plasmid transposon, plasmid transposase; **(b)** cross-section of (a) where the transposon plasmid mass was 4 µg; **(c)** plasmid transposon, NILV transposase; **(d)** cross-section of (c) where the transposon plasmid mass was 2µg; **(e)** NILV transposon, NILV transposase; **(f)** cross-section of (e) where the NILV transposon vector dose was 1.2µg p24. The rate of integration was assessed by the number of G418-resistant colonies formed and is expressed as a percentage of the transduced or transfected cell number assuming 100% plating efficiency.

The results presented in this section suggest that transposition out of a non-integrating lentiviral backbone is detectable by G418-resistant colony assay, and that expression of transposase from a non-integrating lentiviral vector is sufficient to catalyse this reaction at high vector copy numbers. A 2-fold increase in transposon integration was observed at the optimised transposase vector concentration, suggesting that transposition from episomal NILV DNA into chromosomes occurs at a rate similar to background integration under these conditions.

5.4. The Integration Profile of Hybrid Lentivirus-Transposon Vectors

5.4.1. LM-PCR Recovery of Integration Sites from G418-resistant HeLa Cells

In order to confirm transposition from the non-integrated lentiviral vector, integration sites were recovered from transduced cells by ligation-mediated PCR (LM-PCR). HeLa cells were transduced with integrating LNT/tn[SV40-Neo] or non-integrating LNT/tn[SV40-Neo] alone, non-integrating LNT/tn[SV40-Neo] plus LNT/CMV-SB11, or transfected with plasmids pLNT/tn[SV40-Neo] plus pLNT/CMV-SB11 under the previously optimised conditions. After passaging for 3 days without selection, the cells were trypsinised, re-seeded at a 1 in 10 dilution, and incubated with G418 for two weeks. Genomic DNA was extracted and integration sites were recovered by LM-PCR (Figure 4.10). 3' LTR-specific primers were used to amplify integration sites from cells transduced with integrating or non-integrating LNT/tn[SV40-Neo] alone. Transposon-specific primers were used for cells transfected with the two plasmids or transduced with the two non-integrating lentiviral vectors. LM-PCR products were shotgun-cloned into the TopoTA vector and 192 individual clones for each condition were sequenced commercially using the conventional Sanger method (501).

A Blat search of the sequenced products against the UCSC human genome was used to identify junctions between chromosomal DNA and the transposon IR or the viral 3' LTR. In total, 83 unique, mappable integration sites were recovered from cells transfected with the two Sleeping Beauty plasmids (SB-Plasmid sites) and 26 sites were recovered from cells transduced with both Sleeping Beauty NILVs (SB-NILV sites), confirming transposition from non-integrated lentiviral DNA. In addition, 43 sites were isolated from cells transduced with integrating LNT/tn[SV40-Neo] (ILV sites), but no sites were recovered from cells transduced with non-integrating LNT/tn[SV40-Neo] alone (NILV sites).

In order to increase the recovery of integration sites from these samples, the LM-PCR products were sent for 454 pyrosequencing (502) by our collaborator Richard Gabriel

in the laboratory of Manfred Schmidt at the German Cancer Research Center (DKFZ) in Heidelberg, Germany. In this method, LM-PCR products are amplified by PCR using fusion primers to attach short adaptor sequences to the ends of the molecules. The adaptor sequences allow the LM-PCR products to be annealed to 40μm diameter DNA capture beads under conditions in which one LM-PCR molecule anneals to each bead. An oil-water emulsion is then created so that each bead is confined to its own water solution 'microreactor' which is separated from other beads by the surrounding oil. Further PCR amplification and annealing is performed within this emulsion so that each DNA bead carries many copies of the same LM-PCR product. Beads are then loaded onto a plate covered with 50μm wells so that one bead settles into each well. Sequencing is performed by second-strand DNA synthesis of the LM-PCR products. dNTPs are washed over the plate one by one (e.g. first dGTP, then dCTP, etc). In wells where the supplied dNTP corresponds to the next base in the sequence being synthesised, DNA polymerase adds the relevant dNTP to the nascent DNA strand, releasing pyrophosphate as a by-product. This initiates a reaction cascade within the well such that luciferase in the reaction mixture emits light. This light is detected by a camera and the identity of the dNTPs added in each well is recorded in sequence. The high density of wells on the plate and small reaction volumes within each well enable a large number of sequence reads to be performed economically.

In this study, 1234 unique ILV integration sites were recovered of which 976 could be mapped to a unique genomic position by Blat, 185 SB-NILV sites of which 161 could be mapped, and 917 SB-Plasmid sites of which 752 could be mapped. Although NILV-only genomic DNA containing only background integration events was sent for 454 sequencing, it was not possible to recover a set of sites for integration site analysis. Of the sites which could not be mapped to a unique genomic location, the majority were within known repeat elements (Table 5.1).

	Total unique sites	Sites mapped to unique genomic position			Genomic repeat type of sites not mapped to unique genomic position						
		In gene	Not in gene	Total mapped	SINE	LINE	LTR element	DNA transposon	Satellite or simple repeat	No known repeat	Total not mapped
ILV	1234	751 (0.77)	225 (0.23)	976 (1)	134 (0.52)	33 (0.13)	10 (0.04)	3 (0.01)	3 (0.01)	75 (0.29)	258 (1)
SB-NILV	185	86 (0.53)	75 (0.47)	161 (1)	2 (0.08)	2 (0.08)	1 (0.04)	1 (0.04)	1 (0.04)	17 (0.71)	24 (1)
SB-Plasmid	917	321 (0.43)	431 (0.57)	752 (1)	17 (0.10)	60 (0.36)	10 (0.06)	1 (0.01)	5 (0.03)	72 (0.44)	165 (1)

Table 5.1 Hit genomic features at vector integration sites.

Integration sites were recovered from genomic DNA by ligation-mediated PCR and 454 pyrosequencing. Sequences were mapped to the UCSC human genome (version hg18) by Blat search. The proportion of mapped integration sites within RefSeq genes was scored and is expressed as a proportion of the mapped sites in parentheses. Where sites could not be mapped to a unique site, the hit repeat type was identified using RepeatMasker. The proportion of unmapped sites within each repeat type is expressed in parentheses.

The proportion of ILV integration sites within RefSeq genes was 77%, significantly greater than SB-NILV (53%) and SB-Plasmid (43%) sites ($p<0.001$, χ^2 test). The rate of integration within genes was greater for SB-NILV than for SB-Plasmid ($p<0.013$). To estimate the expected rate of random integration into genes, 1000 random integration sites were also generated through bioinformatics by multiplying the total length of the human genome by a randomly generated number between 0 and 1, and converting this value into a (chromosome, strand, bp) coordinate. 34% of random sites were within genes, and all three vector types were found to integrate within genes more frequently than the random site set ($p<0.001$). The SB-Plasmid and ILV proportions are similar to previous reports (503;504).

5.4.2. Fidelity of Vector-Chromosome Junctions

Inspection of the primary sequence data showed that integration occurred with high fidelity at the ILV, SB-NILV, and SB-Plasmid integration sites (Table 5.2). A complete LTR-chromosome junction was found at 1215 of the 1234 ILV sites, with 19 sites showing small deletions from the end of the LTR. 15 of the 19 junctions were missing one LTR base pair and the remaining 4 varied between 3 and 23 missing base pairs. No bases were found to be missing from Sleeping Beauty IRs in either the SB-NILV or SB-Plasmid datasets.

Integration condition	Junction sequence	Chromosomal location
ILV	**AAATCTCTAGCA**GCAGCAGTTCGTGCTGTGACTTCACTTC	1 q25.2
	AAATCTCTAGCAGTCAAGCATAAAAGTTAAAATAATTTAT	2 q37.1
	AAATCTCTAGCAGGTGTCTACCTTGTCAAGCCTATGTTTG	20 q13.2
	AAATCTCTAGCAGTATGGTTATATTTTTCCTGGTCTAAGG	22 q13.33
SB-NILV	**CGACTTCAACTG**TATGTCTACTTTTCTACATTGGTTGGATTT	1 p22.1
	CGACTTCAACTGTATGTATAGCCTTTTACTTGTTTGAGCCTC	8 q13
	CGACTTCAACTGTAATTTAATTAAATACTGCCAACTTCCCTC	9 p25.3
	CGACTTCAACTGTATATATGTGTGTATATATATGTATATATA	14 q25.2
SB-PLASMID	**CGACTTCAACTG**TAGAAAGCATTGGTTGTCTTTCATCTGGG	1 p36.22
	CGACTTCAACTGTATGTACCTATGACAACAAATGTACCATTA	3 q11.2
	CGACTTCAACTGTATCTCAAGTCAGAGTCACTTGACAAATC	9 q22.1
	CGACTTCAACTGTATTACAAAGAGAATACAAGGCAAAAATA	X p22.31

Table 5.2 Sample integration site sequences recovered from G418-resistant HeLa cells by LM-PCR.

Only the beginning of the flanking chromosomal sequence is shown. Underlined, HIV-1 3'-LTR (ILV) or transposon IR (SB-NILV and SB-PLASMID); not underlined, chromosomal DNA. ILV, integrating lentiviral vector; SB-NILV, Sleeping Beauty transposition from non-integrating lentiviral vector; SB-PLASMID, Sleeping Beauty transposition from plasmid.

5.4.3. Primary Sequence at Integration Sites

The preferred primary sequence composition at HIV-1 and Sleeping Beauty integration sites has been well characterised. Sleeping Beauty insertion occurs exclusively at TA-dinucleotides and shows a weak preference for locations where this dinucleotide is flanked by the palindromic sequence 5'-RCAYATATRTGY-3' (505). The transposon is inserted before the underlined TA and this TA is duplicated at both ends of the insertion by host DNA repair proteins (506). HIV-1 integration also shows a weak preference for a palindromic integration site 5'-TNG(GTNAC)CAN-3', where integration occurs before the 5bp sequence indicated by parentheses and this sequence is also duplicated by host repair proteins at both ends of the insertion site (507;508).

In order to investigate the primary sequence preferences of the recovered integration sites, a Visual Basic script was written to download 100bp of primary sequence flanking each integration site from the UCSC human genome server (March 2006 *hg18* version). Sequences were downloaded for both the real and 1000 site random control datasets.

Biases in base composition at the aligned integration sites were analysed using the WebLogo sequence logo tool (509). In a sequence logo, the height assigned to each base represents the amount of information required to describe its frequency at that position (510). For DNA, which contains 4 possible bases at each position, the maximum information required to describe each position is 2 bits (e.g. in binary, A = 00, T = 01, G = 10, C = 11).

The sequence logos generated for the integration sites from this study are shown in Figure 5.8. As previously reported, the primary sequence preference for integrating lentiviral vectors was found to be weak, but a preferred 5'-GTNAN-3' downstream flanking sequence was detectable (Figure 5.8a). For Sleeping Beauty transposition from both non-integrating lentivirus and plasmid, a ubiquitous TA-dinucleotide immediately flanking the integration site was observed, as well as a wider, weaker palindromic flanking sequence, 5'-ANA(TA)TNT-3' (Figure 5.8b,c). As expected, no information content was detected at randomly-generated integration sites. Interestingly, the heights of the stacks indicate that the total information content of Sleeping Beauty sites was much higher than that of integrating lentivirus sites, at least

at the level of primary sequence. This primary sequence constraint may mean that fewer sites within the genome are suitable substrates for Sleeping Beauty insertion. However, with a total haploid genome size of over 3×10^9 base pairs, the number of accessible sites for Sleeping Beauty insertion is still likely to be very large.

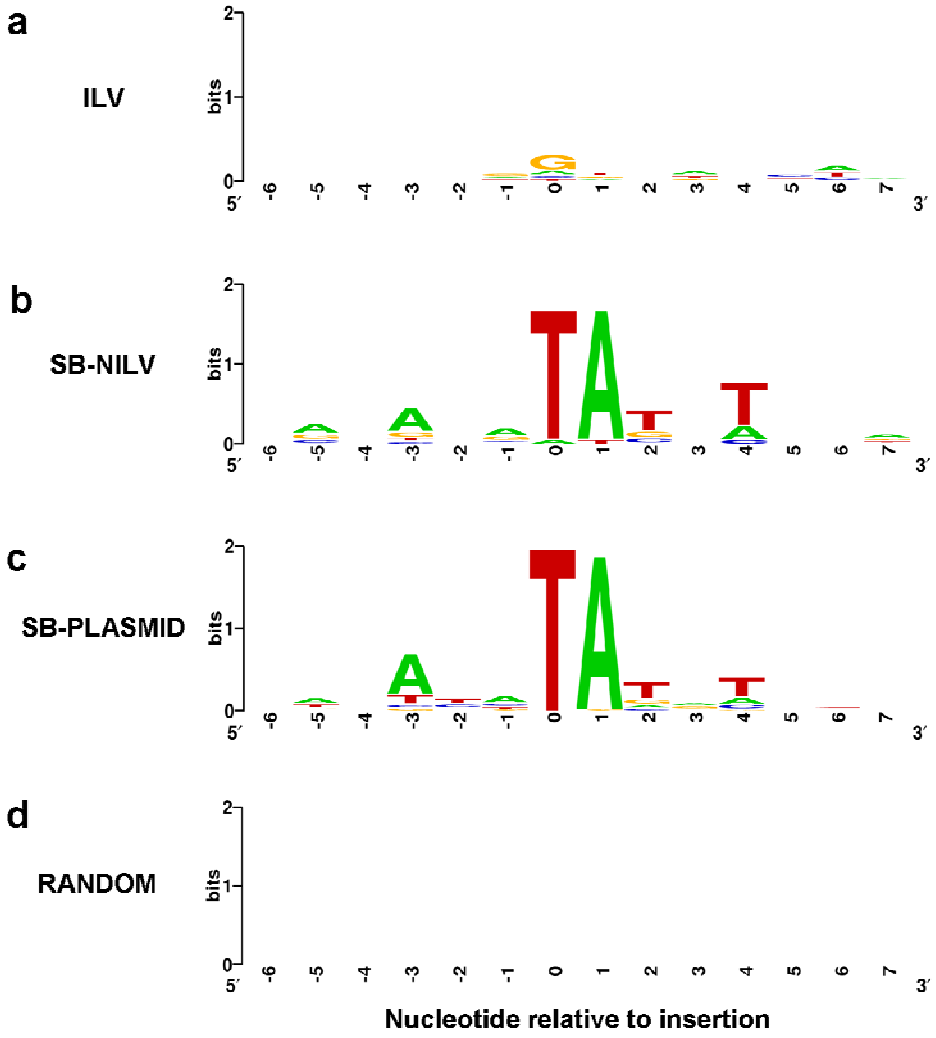

Figure 5.8 Sequence logo for base composition at integration sites.

A sequence logo was generated using the WebLogo tool to identify preferred base-pair usage at the three integration site types. Position 0 denotes the first base of the flanking chromosomal sequence 3' of the integration site. **(a)** ILV, integrating lentiviral vector; **(b)** SB-NILV, transposase-mediated integration of transposons from nonintegrating lentiviral vector; **(c)** SB-PLASMID, transposase-mediated integration of transposons from plasmid; **(d)** RANDOM, randomly generated sites.

5.4.4. Genomic Distribution of Integration Sites

In order to understand the genomic distribution of the three integration site datasets, sites were mapped on to human chromosomes (Figure 5.9) using the UCSC Genome Graphs tool (511). All three datasets contained integrations throughout the genome, as is expected from the known semi-random integration pattern of Sleeping Beauty and HIV-1.

Figure 5.9 Chromosomal map of integration sites.

Sequences were mapped to the UCSC human genome by Blat search and integration sites were depicted relative to chromosomes using the UCSC Genome Graphs tool. Note that HeLa cells are karyotypically abnormal. ILV, integrating lentiviral vector (blue, top line); SB-NILV, transposase-mediated integration of transposons from non-integrating lentiviral vector (red, middle line); SB-PLASMID, transposase-mediated integration of transposons from plasmid (green, bottom line).

To estimate possible biases in integration towards particular chromosomes, the proportion of integration sites on each chromosome was divided by its size relative to the HeLa genome (512) (Figure 5.10a). Integrating lentiviral vectors appeared to show a preference for integration into chromosomes 17 and 19, as has been previously reported (513;514). Biases in Sleeping Beauty integration relative to chromosomes have not been reported in human cells (515;516), but no significant association with particular chromosomes was found in mouse cells (517). Under the assumption that randomly distributed integration sites integrate into chromosomes with a frequency proportional to chromosome size, all three integration site types in this study were found to be non-randomly distributed with respect to chromosomes (ILV $p<0.001$, SB-NILV $p<0.028$, SB-Plasmid $p<0.001$ by χ^2 test).

When integrations within or immediately upstream of genes were mapped relative to their position within the gene or upstream region (Figure 5.10b), no bias towards transcription start sites was detected and no significant variation in integration pattern was observed along the length of the gene. This has been reported for HIV-1 and Sleeping Beauty but not the gammaretrovirus MLV, which integrates preferentially near transcription start sites (518).

It has been reported that lentiviral vector integration occurs preferentially within transcriptionally active genes (519). We estimated the transcription activity of all RefSeq genes in HeLa cells using a published HeLa microarray transcriptome**.** All genes were then categorised as having low, medium, or high transcription activity (each level containing equal numbers of genes). When considering only genes containing integration sites, ILV integration exhibited a clear preference for genes with high levels of transcriptional activity ($p<0.001$), whereas SB-NILV and SB-Plasmid integration showed no significant bias towards any particular level of transcription, and resembled the profile generated for random integration events (Figure 5.10c).

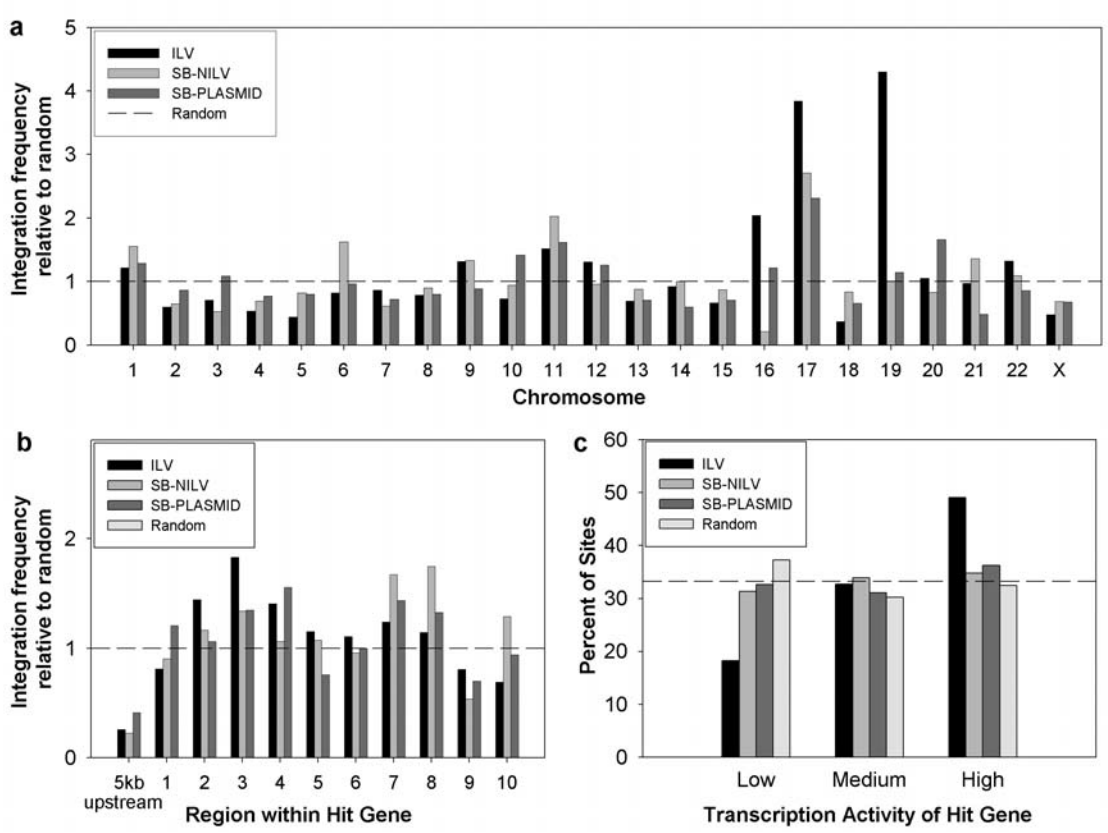

Figure 5.10 Factors influencing integration frequency.

(a) Integration sites per chromosome. The proportion of integration sites on each chromosome was divided by the proportion of the genome accounted for by the chromosome. The dashed line indicates the null hypothesis that the number of integrations per chromosome is proportional to chromosome size. **(b)** Intragenic position of integration sites within genes. RefSeq genes containing integration sites were divided by length into ten equally-sized regions and a 5kb upstream region, and the proportion of integration sites within each region was counted. To allow statistical comparison of integration preferences with average genomic content, 1000 random chromosomal sites were generated by multiplying the total length of the genome by a random number between 0 and 1 and converting this value to a chromosomal coordinate. Vector integration frequencies are expressed relative to the proportion of random sites within each region. **(c)** Transcriptional activity of genes containing integration sites. All RefSeq genes were scored for transcription in HeLa cells using a published microarray dataset (GEO accession number GSM157868). All genes were then assigned to one of three transcription levels (containing equal numbers of genes) to give low, medium, and highly transcribed genes. Integration sites within genes were then scored according to whether the hit gene was transcribed at a low, medium, or high level. For each vector type, the number of intragenic sites per transcription level is expressed as a percentage of the total number of intragenic sites. A dashed line at 33.3% of sites is included to show theoretically equal distribution of sites between the transcription levels. ILV, integrating lentiviral vector, SB-NILV, transposase-mediated integration of transposons from non-integrating lentiviral vector; SB-PLASMID, transposase-mediated integration of transposons from plasmid.

5.5. Summary

This chapter has described the optimisation and characterisation of the hybrid Sleeping Beauty-lentivirus vector:

- A hybrid vector expressing a neomycin phosphotransferase reporter gene was cloned and tested, and a high-throughput G418-resistance integration assay was established.
- The efficiency of Sleeping Beauty transposition per plasmid copy and the level of background integration of plasmid and non-integrating lentiviral DNA were quantified.
- Transposase expression was optimised. Under optimal conditions, the efficiency of transposase-mediated integration was similar to the efficiency of background integration.
- Transposase-mediated integration sites were cloned and sequenced from hybrid vector-transduced cells, demonstrating the feasibility of Sleeping Beauty transposition from non-integrating lentiviral vectors.
- The integration profile of the hybrid vector was found to resemble that of Sleeping Beauty. Hybrid vector integration into active genes was significantly reduced relative to a standard integrating lentiviral vector.

The data presented in this chapter demonstrate that under optimised conditions, Sleeping Beauty transposition from a non-integrating lentiviral DNA can be carried out by transposase expressed from a second non-integrating lentiviral vector. Thus, the concept of a Sleeping Beauty-lentivirus hybrid vector is feasible. Furthermore, the hybrid vector showed a significantly reduced bias for integration into active genes relative to a standard integrating lentiviral vector. The hybrid vector integration profile may therefore be preferable to that of standard integrating lentiviral vectors in terms of the risk of insertional mutagenesis and cell transformation, but this will need to be demonstrated experimentally using established assays for genotoxicity (520;521).

Significant background integration of non-integrating lentiviral vector DNA was also observed in cells transduced with the hybrid vector. The pattern of background

integration, the efficiency of the process *in vivo* rather than cell lines, and the associated risk of insertional mutagenesis are poorly understood and should be studied before non-integrating lentiviral vectors move into clinical trials.

A recent study by Staunstrup *et al* also demonstrated the feasibility of a hybrid Sleeping Beauty-lentivirus vector (522). In the Staunstrup vector, the Sleeping Beauty IRs were inserted into the lentiviral backbone such that transposition would only occur after NILV circularisation into 1- or 2-LTR circles. Around 30% of NILV genomes circularise within three days of transduction (523), and it has been suggested that circular DNA is an efficient substrate for Sleeping Beauty transposition (524). It is not clear whether this is advantageous, as the vector configuration used in this thesis means that the Sleeping Beauty IRs are in the correct orientation for transposition in all NILV species regardless of circularisation state. Furthermore, transposition of the Staunstrup vector after circularisation results in chromosomal integration of lentiviral backbone elements such as the HIV-1 LTRs and the WPRE, while the vector used in this thesis results in transposition of a simpler IR-flanked promoter and transgene cassette.

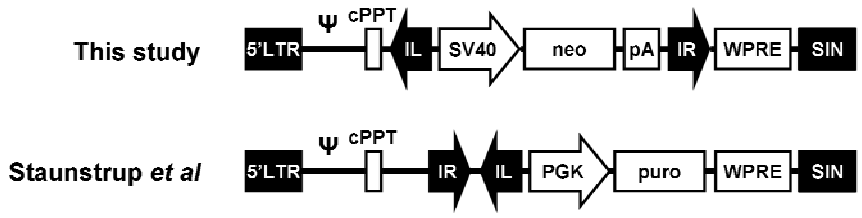

Figure 5.11 Comparison of Sleeping Beauty-lentivirus hybrid vector configurations.
Maps from this study and (525). Both systems use co-transduction with a second non-integrating lentiviral vector for expression of transposase as in Figure 2.11. SV40, simian virus 40 promoter; neo, neomycin phosphotransferase. pA, polyadenylation signal from simian virus 40; 5'LTR, HIV-1 5' long terminal repeat; Ψ, HIV-1 RNA packaging signal; SIN, self-inactivating (U3-deleted) HIV-1 long terminal repeat; cPPT, central polypurine tract; WPRE, woodchuck posttranscriptional regulatory element; IL, Sleeping Beauty transposon left inverted repeat; IR, Sleeping Beauty transposon right inverted repeat; PGK, human phosphoglycerate kinase promoter; puro, puromycin resistance reporter gene.

An interesting question arising from the present study is whether the pattern of integration observed with the hybrid vector is significantly different from that

observed with plasmid-based Sleeping Beauty transposition. Although both of these systems showed no observable integration bias due to transcriptional activity, the hybrid vector integrated into genes significantly more often than the plasmid-based Sleeping Beauty vector (54% versus 43%, $p<0.001$). Standard integrating lentiviruses are believed to be tethered to chromosomes by the host factor LEDGF/p75, and it is thought that preferential chromosomal binding of LEDGF/p75 within genes may cause HIV-1 integration to be biased towards them (526). Similar tethering of non-integrating lentiviral vector DNA could also bias Sleeping Beauty transposition towards genes. Although this is theoretically possible, a significantly greater integration bias of the hybrid vector towards should be demonstrated with other cell types and reporter genes to confirm that it is a robust effect. In the Staunstrup study, 26 of 87 integration sites recovered from transduced cells were within genes (30%). This is close to the proportion which is theoretically expected from random integration, but the sample size is small and SB-Plasmid and ILV controls were not done for comparison.

In terms of the efficiency of the hybrid vector presented in this study, it was found that HeLa cells transduced with a transposon NILV contained a vector copy number of approximately 30 on day 3 at the highest vector doses tested (Figure 5.5), and under optimised transposase conditions the level of transposition as measured by G418-resistant colonies was 1.2% above background (Figure 5.7). A simple ratio of these values gives an estimated transposition frequency of 4 actively expressed chromosomal insertions per 10^4 NILV copies. At this NILV vector dose, background integration resulted in 1.3% of cells becoming G418-resistant, similarly giving an estimated background integration frequency of around 4 per 10^4 NILV copies, similar to previous reports (527). Under optimised conditions, 40µl of transposon NILV and 5µl of transposase NILV were able to produce 1200 G418-resistant HeLa (after accounting for dilution during seeding) and so have a combined integrating titre of 2.7×10^4 TU/ml. By contrast, a standard ILV with an identical backbone gave a G418-resistant HeLa titre of 2.9×10^8 TU/ml, an improvement of four orders of magnitude.

The efficiency of Sleeping Beauty transposition in terms of integrations per intracellular vector has not been previously reported. Yant *et al* created HeLa cell lines in which a single transposon copy was integrated at a chromosomal location so

that it disrupted a G418-resistance cassette (528). Several cell lines were created, and the transposon was integrated at a different chromosomal position in each. When these cell lines were transfected with a plasmid for transposase expression, the rate of transposon excision could be estimated by the rate at which cells became G418-resistant. The authors report that SB10 caused a mean of 1 in 10^5 HeLa cells in each cell line to become G418-resistant, while an enhanced transposase conferred resistance to 1 in 10^4. Assuming that the rate of transposition from transfected plasmids into chromosomes is similar to that from transposons already integrated into chromosomes, the upper bound estimate for the rate of Sleeping Beauty transposition is 10^{-4} integrations per intracellular vector. The true rate may be lower, as excision is only the first half of the transposition process and it has been reported that approximately half of excised transposons fail to reintegrate elsewhere in the genome (529). Furthermore, Sleeping Beauty reintegration occurs preferentially on the same donor molecule (including plasmid donors), a phenomenon known as 'local hopping' (530), meaning that transposition from a plasmid into a chromosome is expected to be less efficient than excision from a chromosome.

The necessity for integration during the retrovirus life cycle means that the efficiency of retroviral integration is not often considered. However, quantification of DNA intermediates during lentiviral vector transduction of cell lines showed that the total vector copy number peaked at 15-20 per cell while the integrated copy number peaked at 1-2, suggesting an efficiency of integration on the order of 10^{-1} integrations per intracellular vector genome (531).

In their hybrid vector, Staunstrup *et al* used the recently developed hyperactive Sleeping Beauty transposase SB100 which has been reported to be 100-fold more active than the original SB10 (or approximately 30-fold more active than SB11) (532;533). In that study, the rate of hybrid vector integration was 25-fold greater than background integration when transposase was expressed from a plasmid in target cells, or 10-fold greater than background when transposase was expressed from a second non-integrating lentiviral vector. This compares to a 2-fold increase in integration relative to background in the present study. These data suggest that even using the most active transposase available, the integrating titre of a hybrid Sleeping Beauty-lentivirus vector is around 3 orders of magnitude less than that of standard

ILVs, which integrate 10^4-fold above background as measured by NILV integration (534). It should be noted here that when Staunstrup *et al* performed a transduction with equal p24 doses of either an ILV or their hybrid vector, only 12-fold fewer puromycin-resistant colonies resulted from the hybrid condition. However, neither the p24 vector dose nor the number of cells seeded for colony counting were detailed for this result, and it is difficult to reconcile it with the study's reported integration efficiencies relative to background integration.

It is possible that all presently available Sleeping Beauty transposases are catalytically much less active than HIV-1 integrase. HIV-1 integrase and the Sleeping Beauty transposase are quite similar, evolutionarily related proteins which catalyse the integration reaction using essentially the same nucleophilic attack mechanism via a DDE catalytic domain (535), and it is not immediately clear why such similar proteins should be so different in their catalytic activities. A second possibility is that HIV-1 integrase has a kinetic advantage in that the formation of the preintegration complex takes place using co-localised components within the virion and so is not rate-limiting for integration, whereas the Sleeping Beauty transposase must be expressed in the cell, bind to the transposon, catalyse its excision and the formation of the preintegration complex, and only then catalyse the integration reaction. By performing the early steps of this process in a more conducive environment within the virion, HIV-1 integrase may obtain a large efficiency gain.

An informative test might be an *in vitro* integration assay using purified HIV-1 integrase and Sleeping Beauty transposase to estimate the enzyme catalytic coefficient (536), but such assays are technically challenging and perhaps not capable of yielding quantitative comparisons. However, it is feasible to clone the Sleeping Beauty transposase into the HIV-1 Gag-Pol polyprotein in a manner that mimics HIV-1 integrase's own incorporation, so that Sleeping Beauty transposase would be transported within virions as a protein. Perhaps more importantly, this approach has the advantage that no DNA expressing the transposase protein would need to enter the target cell, which is preferable as it reduces the risk of background integration of the transposase gene into target cell chromosomes. Research towards this approach is described in the next chapter.

6. Delivery of Transposase Protein by Lentiviral Virions

6.1. Aims

- To construct fusion proteins which allow packaging of eGFP and transposase into lentiviral virions
- To investigate virion incorporation and processing of these proteins
- To determine the effect of heterologous protein incorporation on virion function
- To assess delivery of eGFP and transposase to target cells by lentiviral virions

6.2. Introduction

Protein transduction is direct delivery of proteins to target cells, and has been an active area of study since it was first reported that the Tat protein of HIV-1 contains an 11 amino acid protein transduction domain (PTD) which can be fused to heterologous proteins, allowing them to enter cultured cells directly from the surrounding medium (537-539). PTDs able to deliver heterologous proteins were subsequently identified in other proteins, including the *Drosophila* Antennapedia homeodomain protein (540) and Kaposi fibroblast growth factor (541). The mechanism of protein transduction is controversial, but it has been suggested that proteins fused to highly cationic peptides such as the Tat-PTD first interact with negatively charged cell surface molecules such as proteoglycans, and that these bound proteins undergo macropinocytosis before escape into the cytoplasm (542). Tat-mediated protein transduction is being developed for a number of therapeutic applications, including treatments for cancer, ischemia, and neurodegenerative diseases (reviewed in (543)).

The ability to manipulate the protein content of cells without gene transfer may be advantageous in removing the risk of insertional mutagenesis due to integration of transgenic DNA. In addition, protein transduction of enzymes with the ability to rearrange DNA such as Cre recombinase or Sleeping Beauty transposase would ensure that their presence is transient in both dividing and non-dividing target cells,

reducing potential genotoxic risks associated with these proteins. Successful protein transduction of enzymes with the ability to rearrange DNA has been shown previously, such as lipid-mediated transfection of the AAV Rep protein to mediate site-specific integration in target cells (544) or protein transduction using a Cre recombinase fused to a cell-penetrating peptide (545).

In a hybrid vector such as the Sleeping Beauty-lentivirus vector presented in previous chapters, it may be advantageous to deliver the integrating protein as an internal component of the gene transfer vector (Figure 6.1). This approach mimics the behaviour of retroviruses, which integrate using molecules of retroviral integrase which enter target cells in a complex with the viral genome and other components. Co-delivery of the integrating protein in this way may be a more efficient strategy than expression of the integrase gene within target cells, for example by removing the need for the integrating protein to locate the viral genome within the cell or by minimising the delay between cell entry and integration of the viral genome, during which time the viral genome may be lost to processes within the cell.

HIV-1-mediated delivery of heterologous proteins has been previously reported. The general approach is to fuse the heterologous protein to an HIV-1 protein so that it is incorporated into vector particles in producer cells, and then to transduce target cells and investigate the intracellular distribution of protein. Virion-mediated protein transduction may enable protein delivery with the broad cell tropism of VSVg-pseudotyped lentiviral vectors. Fusions are most commonly made with the Vpr protein, such as a Vpr-eGFP fusion to investigate the biology of HIV-1 assembly (546) or cell entry and trafficking (547;548), a Vpr-β-lactamase fusion to measure virion fusion efficiency (549), and a Vpr-linamarase fusion for prodrug-mediated tumour killing (550). Vpr may be a good candidate for protein transduction fusions as it remains associated with the viral DNA in the cytoplasm, thus demonstrating its ability to escape from endosomes (551). A Nef-HSV-TK fusion for prodrug-mediated cell killing has also been reported (552).

Although these authors were able to show protein transduction using HIV-1, alternative vector configurations may be preferable. Firstly, accessory proteins such as Vpr and Nef have been removed from second and third generation lentiviral vectors in order to simplify vector composition (553) and eliminate undesirable effects these

accessory proteins may have on target cells (554;555). Secondly, the level of accessory protein incorporation into virions is not high, with a 1:7 Vpr:Gag and 1:20 Nef:Gag molar ratio in each virion. (There are estimated to be 2000-5000 molecules of Gag in each HIV-1 virion (556).) Thus, the stoichiometry of HIV-1 virions means that direct fusion to Gag has the largest potential cargo capacity for virion-mediated protein transduction.

This chapter presents work towards the development of HIV-1-mediated protein transduction to deliver the Sleeping Beauty transposase.

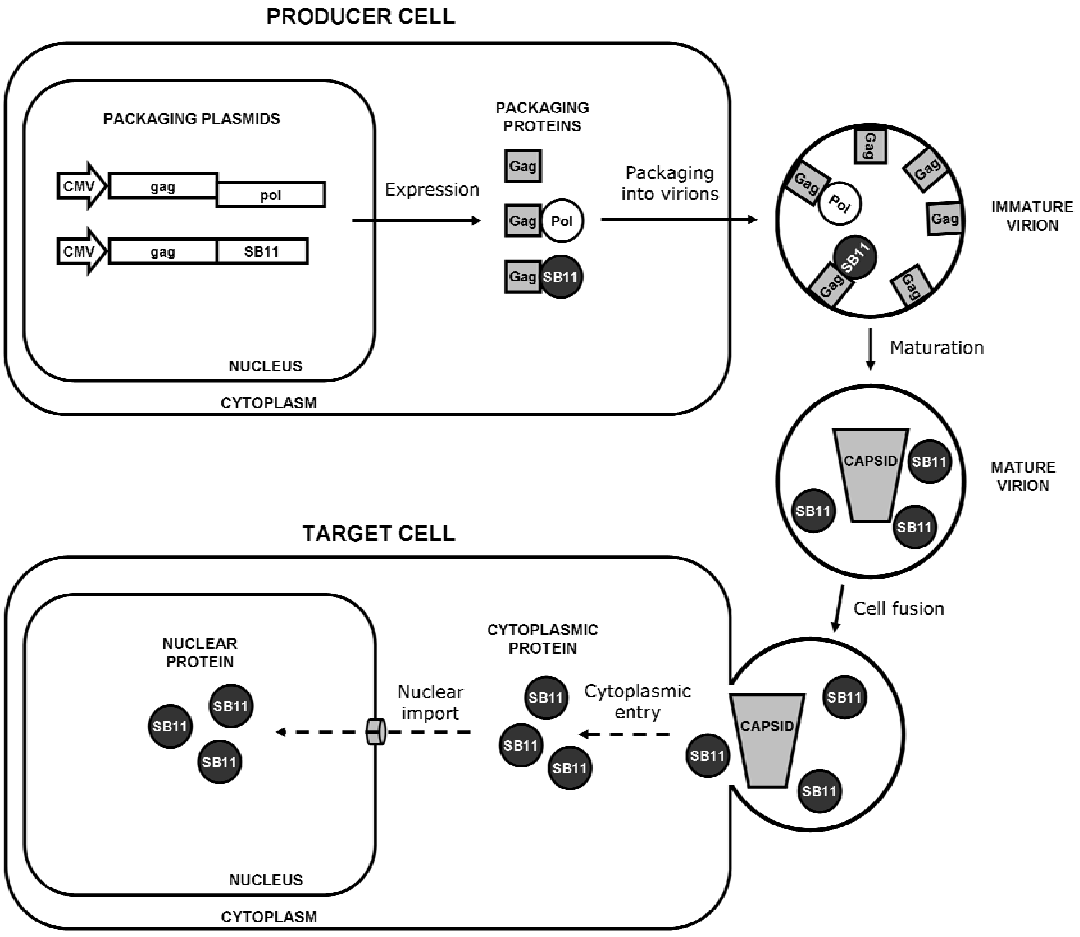

Figure 6.1 Schematic overview illustrating the goal of HIV-1-mediated protein transduction.
Sleeping Beauty transposase is expressed in the producer cell as a fusion with the HIV-1 Gag protein, allowing it to be packaged into virions. During virion maturation, HIV-1 protease cleaves the protease cleavage site between Gag and transposase. When the mature virion fuses with the target cell, transposase is released into the cytoplasm at the same time as the viral capsid. The nuclear localisation signal present within transposase may enable nuclear import of the protein.

6.3. Design and Plasmid Subcloning of Gag-eGFP and Gag-Transposase Fusion Proteins

Gag is present in virions at the highest stoichiometric ratio of any protein in the virion lumen. Since this offers the highest potential protein transduction cargo capacity, constructs were generated in which eGFP and SB11 were fused to Gag. The goal was to generate fusion proteins which would not interfere with normal virion assembly, maturation, and infection. Sleeping Beauty transposase tolerates N- and C-terminal fusions poorly (557;558), so constructs were generated in which Gag and eGFP or SB11 were separated by a pre-existing protease cleavage site (Figure 6.2).

In order to minimise the effects of heterologous protein fusion on Gag multimerisation during virion assembly, proteins were fused near the C-terminus of Gag, away from the Gag I-domain which is necessary for Gag-Gag interactions (559).

The HIV-1 Gag and Gag-Pol proteins are expressed and packaged in the producer cell as polyproteins. During virion maturation, HIV-1 protease undertakes sequential cleavage of the polyproteins at the protease cleavage sites which link their constituent proteins. Processing of the polyproteins takes place in an ordered fashion due to differences in the rate of cleavage at each site. Ordering is regulated by polyprotein interactions with RNA, local sequence, the presence of spacer peptides, and possibly oligomerisation of proteolytic intermediates (560-564). In HIV-1 Gag, cleavage occurs first between the p2 peptide and NC, then at the MA-CA and p1-p6 boundaries, and finally at the CA-p2 and NC-p1 boundaries. In Gag-Pol, p2-NC cleavage is followed by cleavage at the RT-IN and MA-CA boundaries, and lastly sites flanking PR (565). It is not known whether the ordering of cleavage events has any biological significance. In order to ensure proper proteolytic processing of eGFP and SB11 from the fusion proteins, the p6 protein of Gag was replaced with these proteins while retaining the p1-p6 cleavage site. The first 5 amino acids of p6 were retained to ensure the proper context for proteolytic processing (see Appendix 8.1 for annotated amino acid sequences).

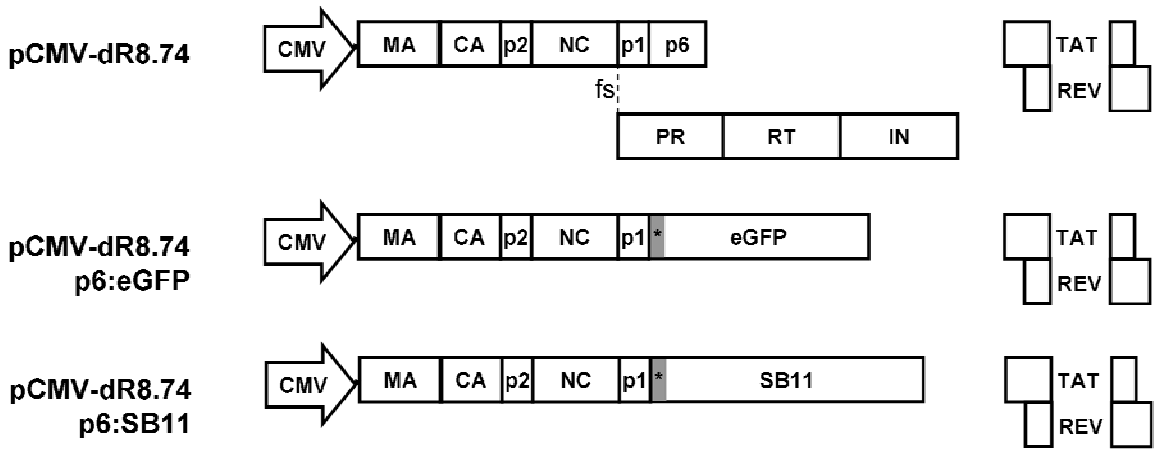

Figure 6.2 Protein transduction constructs.

The standard second generation lentiviral vector packaging plasmid is pCMV-dR8.74, while pCMV-dR8.74 p6:eGFP and pCMV-dR8.74 p6:SB11 are the fusion constructs generated in this study. CMV, human cytomegalovirus immediate early promoter; MA, matrix protein; CA, capsid protein; p2, linker peptide; NC, nucleocapsid protein; p1, linker peptide; p6, Vpr-interacting protein; TAT, transactivator of transcription protein; REV, enhancer of translation; fs, ribosomal frameshift site; PR, protease; RT, reverse transcriptase; IN, integrase; eGFP, enhanced green fluorescent protein; SB11, Sleeping Beauty transposase; *, first 5 amino acids of p6 fused to the N-terminus of eGFP or SB11.

To generate the SB11 fusion construct, pBluescript was digested with SacII, blunted using T4 DNA polymerase, and then religated to destroy the site. A fragment of the pCMV-dR8.74 *gag* sequence was amplified by PCR using Pfu polymerase and primers FWD 5' – TAA TCA GAATTC GGC CGG CCG CGT TG and REV 5' – ATA ACT GAATTC CCG CGG GGT CTG CTC and cloned into this plasmid using EcoRI. SB11 was digested from pCMV-SB11 using SacII and cloned into this plasmid at the SacII site so that the transposase and viral coding sequences were in the same orientation. The *gag*:SB11 fusion fragment was cloned into pCMV-dR8.74 between the ApaI and EcoRI sites so that it replaced the p6 and *pol* coding sequences.

To generate the eGFP fusion construct, eGFP was amplified from LNT/SFFV-eGFP using primers FWD 5' CAG TCC GCG GCC TGC AGG GTA TGG TGA GCA AGG GCG AG and REV 5' CGC ACC CGC GCC CGG TCG ACT TAC TTG TAC AGC TCG TC and this fragment was cloned in place of SB11 in the

pBluescript/*gag:*SB11 plasmid using SacII and so that eGFP and *gag* were in the same orientation. The *gag*:eGFP fragment was cloned into pCMV-dR8.74 between the ApaI and SalI sites.

6.4. Expression, Virion Incorporation, and Processing of Fusion Proteins

6.4.1. Expression and Membrane Targeting of Gag-eGFP

To test expression of the Gag-eGFP fusion protein, 293T cells were transfected with either the standard non-integrating lentiviral packaging plasmid or the Gag-eGFP fusion plasmid (Figure 6.3). Green fluorescence was clearly visible, indicating expression of the fusion protein.

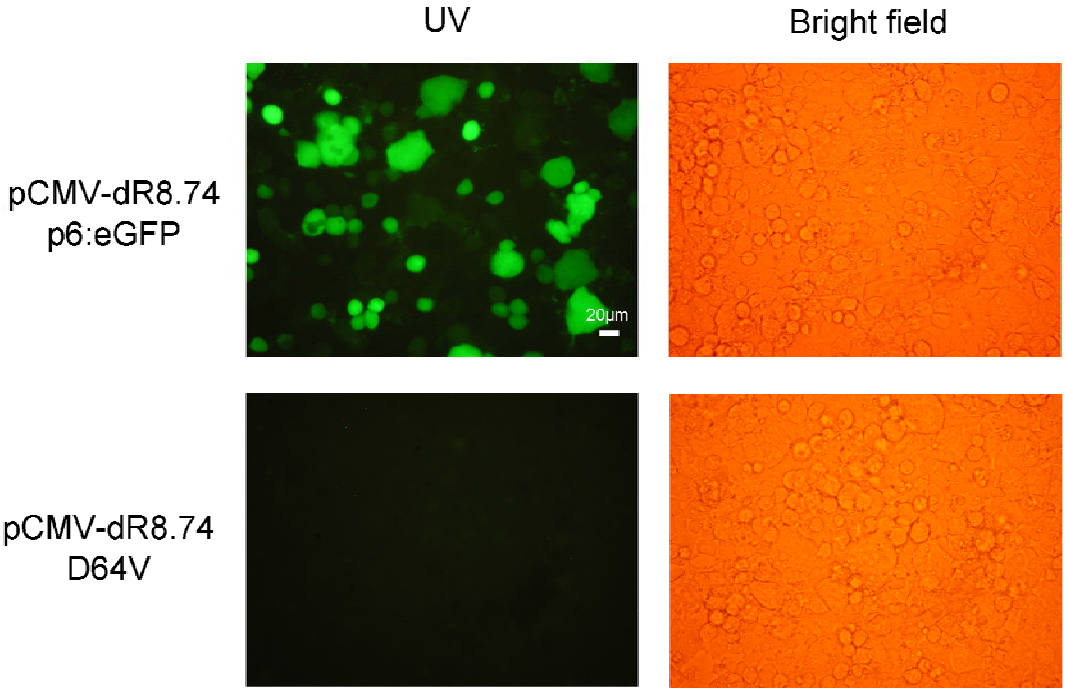

Figure 6.3 Fluorescence microscopy to demonstrate expression of Gag-eGFP.
10^5 293T cells were transfected with 800ng pCMV-dR8.74 D64V or pCMV-dR8.74 p6:eGFP. eGFP expression was imaged by fluorescence microscopy using a 40x objective 48 hours post-transfection.

Gag is myristoylated during translation in the cytosol (566), whereby the 14-carbon fatty acid myristate is attached to the N-terminal glycine of Gag. The myristate moiety acts together with a basic N-terminal membrane binding (M) domain (567) to allow Gag to become membrane-associated.

In order to test the hypothesis that the Gag-eGFP fusion protein is membrane-associated in the same way in producer cells, 293T cells were transfected with

pCMV-dR8.74 p6:eGFP and pLNT/SFFV-eGFP and imaged by confocal microscopy (Figure 6.4). In cells expressing the Gag-eGFP fusion protein, green fluorescence was seen preferentially at the plasma membrane, while cells expressing the standard eGFP were fluorescent throughout the cell volume. These results suggest that the Gag-eGFP fusion protein is expressed and targeted to the plasma membrane correctly.

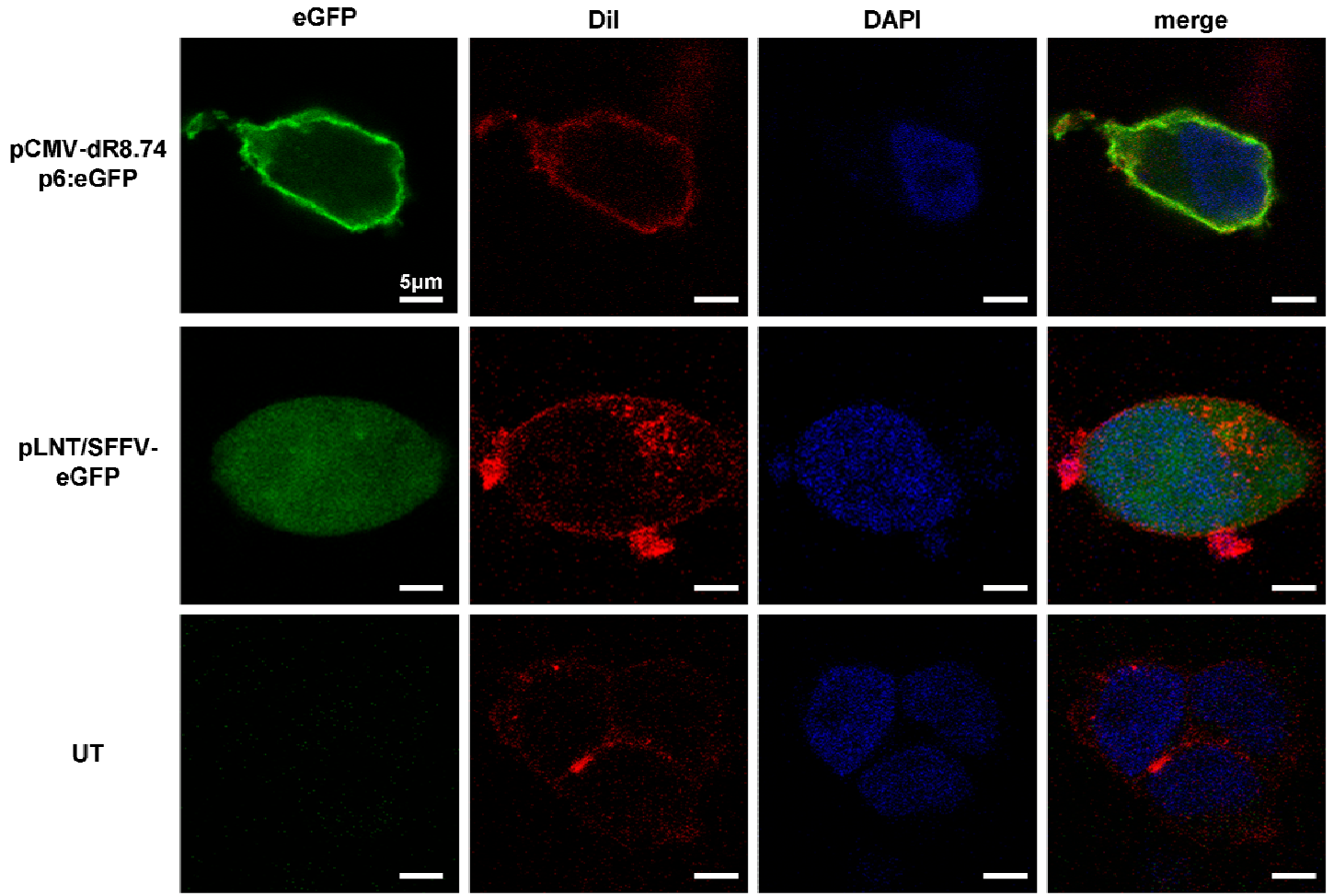

Figure 6.4 Confocal microscopy to show subcellular localisation of Gag-eGFP in producer cells.

10^5 293T cells were transfected with 800ng pCMV-dR8.74 p6:eGFP or pLNT/SFFV-eGFP. 48 hours post-transfection, cells were stained with the lipophilic die DiI to allow visualisation of the plasma membrane and the nuclear stain DAPI. Images represent an xy slice with a 1µm z-axis thickness and were captured with a 63x/1.4 oil objective lens. UT, untransfected.

6.4.2. Production and Titration of Lentiviral Vectors Containing Gag Fusion Proteins

In order to assess the tolerance of lentiviral vector assembly and budding to co-expression of Gag-eGFP and Gag-SB11, vector was produced in the standard way (293T transfection with PEI, VSVg pseudotype, concentration by ultracentrifugation) except that the mass of packaging plasmid (37.5µg per $12x10^6$ 293T cells) was held constant and divided between the standard pCMV-dR8.74 D64V Gag-Pol and the fusion Gag-eGFP or Gag-SB11 plasmids in fixed ratios. The ratios tested were 1:0 Gag-Pol:Gag-Fusion, 30:1, 10:1, 3:1, 1:1, 1:3, 1:10, 1:30, and 0:1, so that a 1:0 ratio represents standard non-integrating plasmid only while a 0:1 ratio represents vector containing the fusion protein only and no full-length Gag-Pol. Vector was produced simultaneously in parallel for all ratios described and titred in duplicate by p24 ELISA (Figure 6.5).

The vector titre appeared to be unaffected by increasing the proportion of fusion protein up to a plasmid ratio of 1:3 Gag-Pol:Gag-eGFP or Gag-Pol:Gag-SB11, beyond which the physical titre declined. This suggests that virion assembly and budding is able to tolerate a high level of fusion protein co-expression, though it is important to note that high expression of fusion protein within producer cells does not necessarily imply that high levels of fusion protein are incorporated into virions that bud away from these cells.

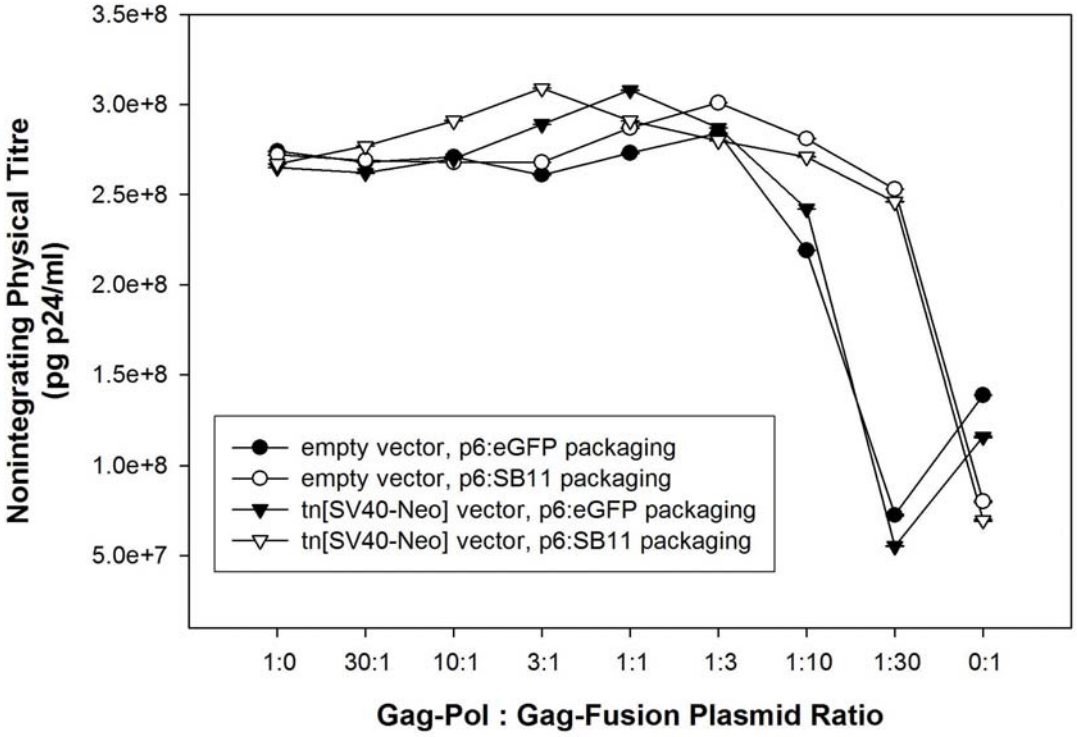

Figure 6.5 Titration of lentiviral vector containing Gag-Fusion constructs by p24 ELISA.

Vector was generated in parallel as described in the text and titred in duplicate. Empty vector, lentivirus vector genome containing LTRs, packaging signal, cPPT, and WPRE but no promoter-transgene cassette; tn[SV-40-Neo] vector, lentivirus-transposon vector genome.

6.4.3. Western Blotting of Vectors for Protein Transduction

Gag and Gag-Pol are initially expressed and packaged into virions as polyproteins. During or soon after virion budding away from the producer cell, maturation occurs whereby the HIV-1 protease sequentially cleaves these polyproteins into their constituent proteins, allowing assembly of the mature virion core and causing the virion to become infectious.

The fusion constructs described in this chapter were designed to retain the p1-p6 protease cleavage site N-terminal to eGFP or SB11. To assess proteolytic processing of fusion proteins incorporated into virions, samples of concentrated vector were lysed under denaturing conditions and the component proteins were separated by electrophoresis and visualised by Western blotting (Figure 6.6).

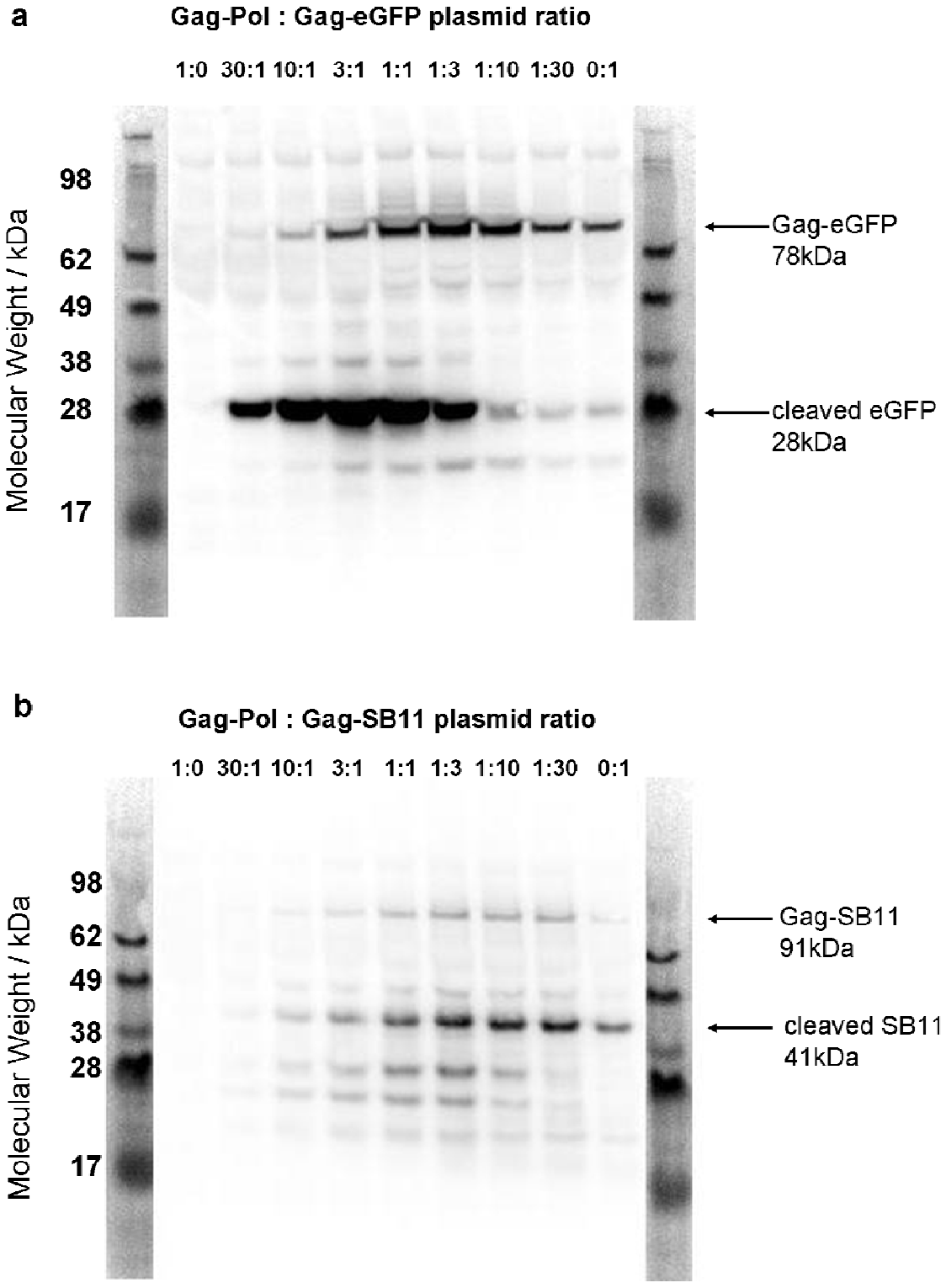

Figure 6.6 Western blot to show proteolytic processing of Gag-Fusion proteins.

Lentiviral vectors were prepared by transfection of 293T cells as described in Section 1.1.1 and concentrated by ultracentrifugation. 10µl of each preparation was lysed in reducing Western lysis buffer and separated by polyacrylamide gel electrophoresis followed by Western blotting. **(a)** Vector

containing eGFP fusion proteins was probed with mouse anti-eGFP primary antibody; **(b)** Vector containing SB11 fusion proteins was probed with mouse anti-SB11 transposase primary antibody.

When vector containing the Gag-eGFP polyprotein was produced, both unprocessed Gag-eGFP and fully cleaved eGFP forms were visualised (Figure 6.6a). The concentration of fully cleaved eGFP in the vector preparation increased with the proportion of Gag-eGFP plasmid in the producer cell up to a 3:1 Gag-Pol:Gag-eGFP plasmid ratio, and declined thereafter. The concentration of unprocessed Gag-eGFP peaked at a Gag-Pol:Gag-eGFP plasmid ratio of 1:3.

Vector containing Gag-SB11 fusion protein showed a similar but less clear trend on Western blotting (Figure 6.6b). The concentration of fully processed SB11 protein peaked at a higher proportion of Gag-SB11 plasmid (1:3 Gag-Pol:Gag-SB11), implying that the Gag-SB11 protein is not expressed and/or packaged as efficiently as Gag-eGFP. In addition, multiple smaller bands were detected whose intensity is proportional to that of the full-size SB11 bands in their respective lanes, indicating that they are truncated forms of SB11. It has been previously reported that that the Sleeping Beauty transposase is extremely prone to folding errors and truncation (568).

In summary, Western blotting suggested that both eGFP and SB11 were packaged into virions and could be correctly processed by the HIV-1 protease.

6.5. Protein Delivery to Target Cells

6.5.1. Visualisation of Protein Delivery by Microscopy

The objective of protein transduction is to deliver functional heterologous proteins to target cells. As proof of this principle, eGFP fluorescence was used to assess delivery to cells by fusion protein vectors. 293T cells were transduced with concentrated vectors containing varying ratios of the Gag-eGFP fusion protein and visualised by UV microscopy (Figure 6.7).

Green fluorescence was clearly visible in wells transduced with vestor prepared with Gag-Pol:Gag-eGFP ratios between 3:1 and 1:10. Little fluorescence was observed at more extreme Gag-Pol:Gag-eGFP ratios. It is important to note that this assay would not distinguish between functional virions containing Gag-eGFP protein and cell debris containing eGFP which was co-purified during lentiviral vector preparation.

To assess whether the green fluorescent matter present within the well was taken up by target cells, 293T cells were seeded onto polylyisine-treated microscope slides, transduced as above, and visualised by confocal microscopy (Figure 6.8).

Punctate eGFP fluorescence was clearly visible at the cell surface and within cells for all vectors containing Gag-eGFP. This eGFP distribution was significantly different from that in cells transduced with an equal volume of integrating eGFP lentiviral vector (prepared in parallel), which showed homogenous eGFP fluorescence throughout the cytoplasm and nucleus.

The punctate fluorescence observed could mean that material containing eGFP is sequestered in a membrane-bound organelle, possibly endosomes. It is important to note that punctate fluorescence has been previously observed with Tat-eGFP protein transduction, but that this pathway is able to deliver useful quantities of therapeutic proteins (569). Again, it should be emphasised that this assay does not demonstrate that the green fluorescent material taken up by target cells represents functional lentiviral virions containing eGFP.

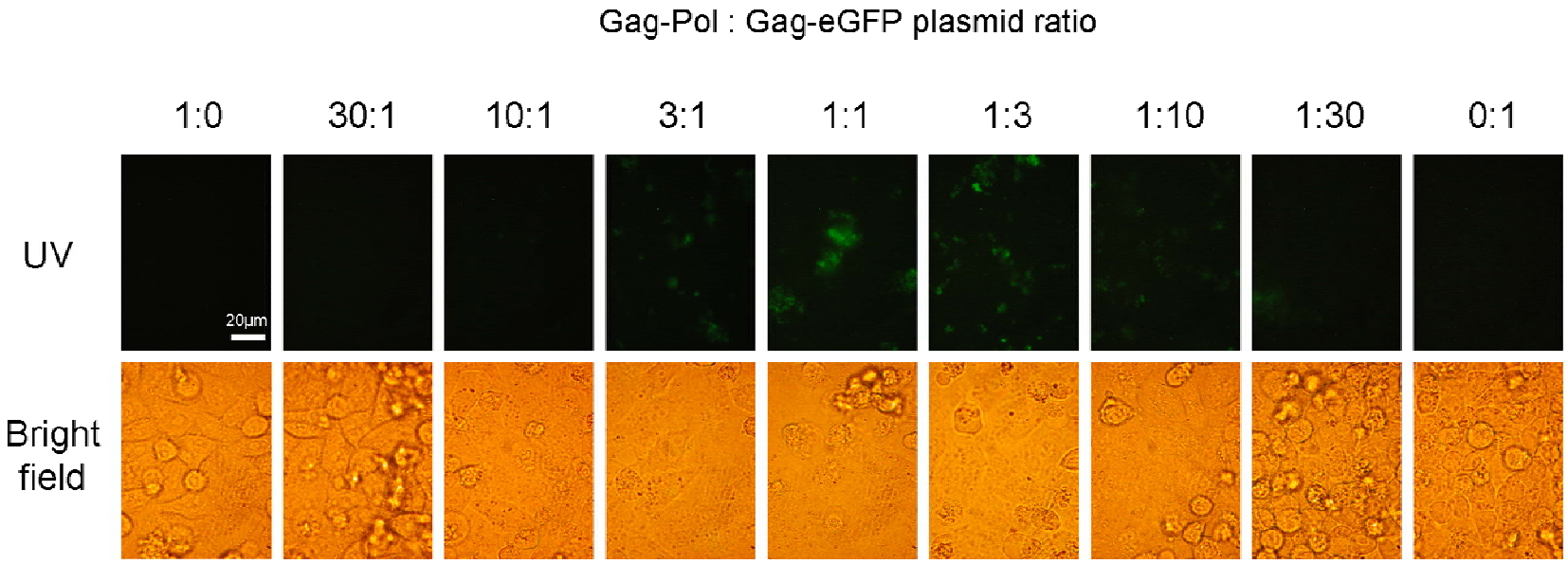

Figure 6.7 Fluorescence microscopy to visualise Gag-eGFP virus transduction.

10^5 293T cells were transduced with concentrated lentiviral vector (diluted 8-fold in culture medium) produced with the Gag-Pol : Gag-eGFP plasmid ratios indicated. 6 hours post-transduction, cells were imaged by bright field and UV (ultraviolet) microscopy using a 40x objective.

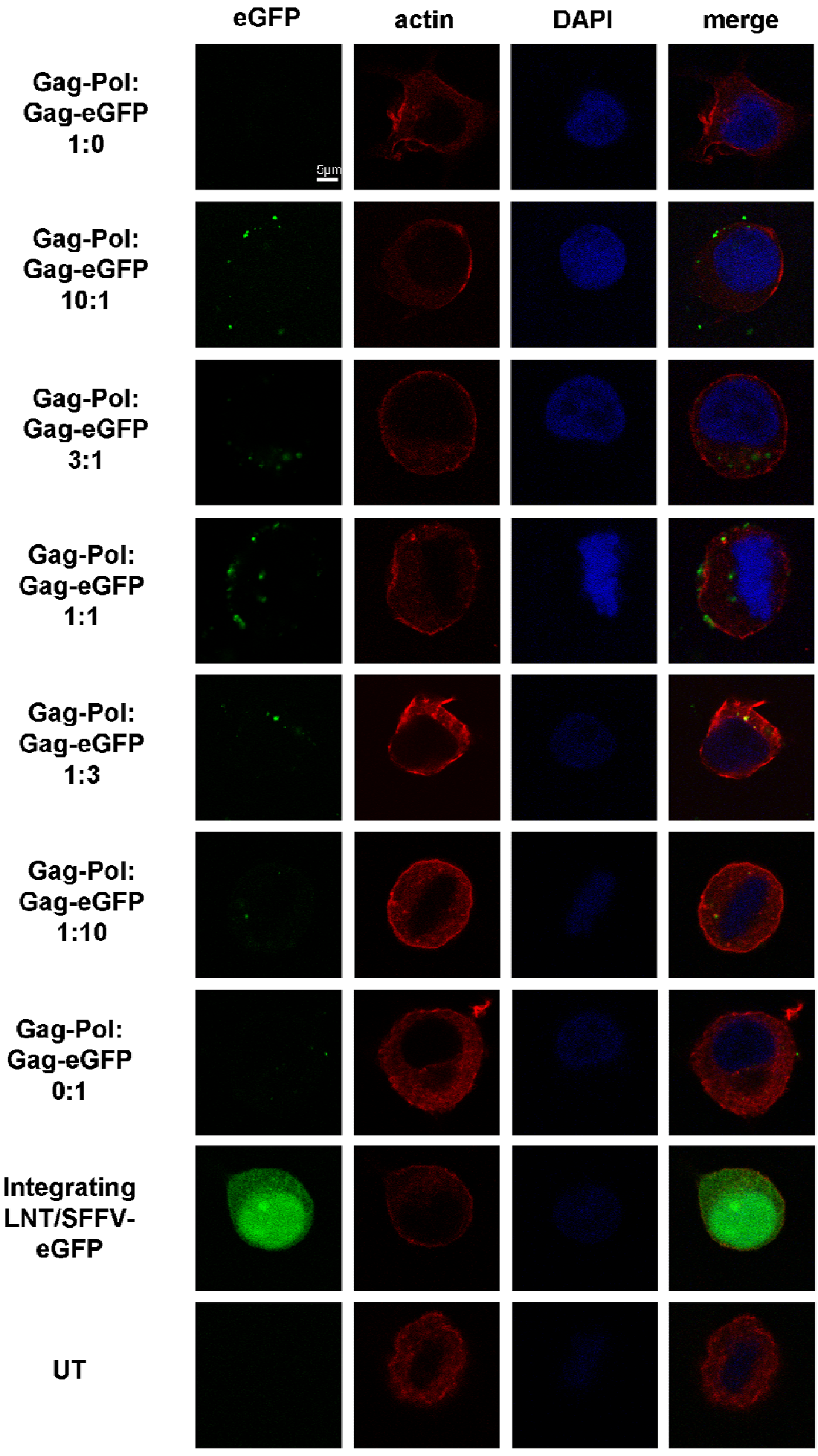
eGFP
actin
DAPI
merge
Gag-Pol:
Gag-eGFP
1:0
5µm
Gag-Pol:
Gag-eGFP
10:1
Gag-Pol:
Gag-eGFP
3:1
Gag-Pol:
Gag-eGFP
1:1
Gag-Pol:
Gag-eGFP
1:3
Gag-Pol:
Gag-eGFP
1:10
Gag-Pol:
Gag-eGFP
0:1
Integrating
LNT/SFFV-
eGFP
UT

Figure 6.8 Confocal microscopy to assess subcellular eGFP distribution following lentiviral protein transduction.

10^5 293T cells were transduced with concentrated lentiviral vector preparation (diluted 8-fold in culture medium) produced with the indicated Gag-Pol : Gag-eGFP plasmid ratios. 24 hours post-transduction, cells were stained with rhodamine-phalloidin to allow visualisation of the actin cytoskeleton and the nuclear stain DAPI. Images represent an xy slice with a 1μm z-axis thickness and were captured using a 63x/1.4 oil objective lens. UT, untransduced.

6.5.2. Measurement of Protein Delivery by Integration Assay

An objective of the work presented in this chapter was to enable transposase protein incorporation into virions together with the transposon vector genome. It was hypothesised that co-delivery of these two elements would result in transposase-mediated integration (Figure 6.1).

To test the feasibility of the system, fusion protein vector was produced in parallel for both Gag-eGFP and Gag-SB11. As before, the plasmid ratio of Gag-Pol to Gag-eGFP or Gag:SB11 was varied to allow optimisation of the balance between vector fitness and heterologous protein incorporation. The vector genome was pLNT/tn[SV40-Neo].

HeLa cells were transduced with a range of vector concentrations and the integration efficiency was determined by G418-resistant colony formation (Figure 6.9). It was found that increasing the proportion of SB11 transposase in virions did not result in a detectable increase in the level of integration above background as measured by transduction with the Gag-Pol-only vector. At each Gag-Pol:Gag-SB11 ratio, the rate of integration was not significantly different to that observed in the matching Gag-Pol:Gag-eGFP condition (Figure 6.9c). In addition, increasing the proportion of Gag-SB11 plasmid beyond 3:1 Gag-Pol:Gag-SB11 or Gag-eGFP resulted in greatly reduced background integration, implying a significant loss of vector fitness at these ratios.

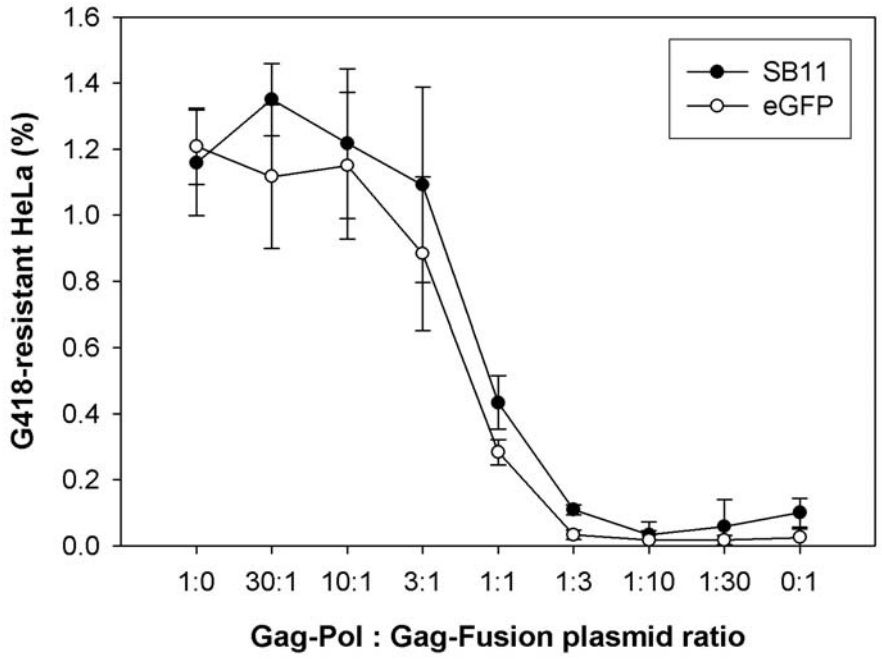

Figure 6.9 Integration assay to assess the effect of transposase fusion protein incorporation on the level of integration.

Concentrated lentiviral vectors were produced in parallel containing either Gag-eGFP or Gag-SB11 and the pLNT/tn[SV40-Neo] vector genome. 10^5 HeLa cells were transduced with an 8-fold dilution of these vectors, incubated for 3 days without selection, and then re-seeded into fresh 24-well plates at a 1:50 dilution and incubated in the presence of 1mg/ml G418 for 14 days. Integration efficiency was assessed by G418-resistant colony counting.

6.6. Summary

This chapter described the construction and development of the Gag-eGFP and Gag-SB11 fusion proteins to allow lentiviral protein transduction of target cells:

- Fusion proteins were cloned and Gag-eGFP was found to express in producer cells
- Fusion to Gag was found to target eGFP to the plasma membrane in producer cells
- Proteolytic processing of fusion proteins was demonstrated by Western blotting concentrated lentiviral vector preparation
- Incorporation of transposase protein into virions did not result in a detectable increase in the level of vector integration. The fitness of vector particles was found to decline where the Gag-Pol:Gag-SB11 or Gag-eGFP ratio was less than 3:1.

The data presented in this chapter demonstrate that it is possible to incorporate heterologous proteins into lentiviral virions when fused to Gag, and these fusion proteins can be correctly processed by the HIV-1 protease during virion maturation to release the heterologous protein.

When cells were transduced with lentiviral vector preparation produced with the Gag-eGFP fusion protein, green fluorescent matter was detectable within cells by confocal microscopy 24 hours post-transduction. However, significant further study would be required to determine whether this material represents functional lentiviral particles or simply debris containing eGFP which was co-purified with virions during vector preparation.

In this study, transposition from an NILV genome catalysed by transposase protein transduction was not observed. This could be due to a number of factors, including an absence of transposase within functional virions even using this targeting strategy.

7. Discussion

The growing number of gene therapy clinical trials resulting in a clear therapeutic benefit for patients has underscored the power of this approach to treat conditions for which conventional alternatives are unattractive or unavailable (32;570-576). Many of these trials have used gene transfer vectors which integrate their DNA into patient chromosomes, allowing therapeutic DNA to be passed to daughter cells during mitosis and enabling stable gene expression in dividing tissues. However, five serious adverse events which occurred in two clinical trials for X-linked severe combined immunodeficiency have raised concerns as to the safety of commonly used integrating vectors (577;578). In these patients, semi-random integration of gammaretroviral vectors into patient chromosomes appears to have dysregulated expression of nearby T-cell proto-oncogenes such as LMO2, leading to T-cell leukaemia-like expansions (579).

It has been suggested that the genomic integration pattern of a vector may influence its ability to cause insertional mutagenesis in this way. The integration preferences of the major vectors have now been well-characterised. Gammaretroviral vectors integrate preferentially near transcription start sites, while lentiviral vectors show a preference for integration within actively transcribed genes (580;581). Adeno-associated virus vectors integrate preferentially within CpG islands and the first 1kb of genes (582). The Sleeping Beauty transposon has a near random integration profile with a low preference for genes and no detectable bias towards regions of active transcription (583;584).

An attractive approach to addressing the problem of insertional mutagenesis may be to alter the genomic integration pattern of gene transfer vectors in such a way as to reduce the probability of oncogenic events such as proto-oncogene activation or tumour suppressor gene inactivation. This could be achieved through the development of hybrid vectors which combine the gene transfer activity of one vector with the preferable integration profile of another.

In this study, a hybrid vector was developed which combines the efficient gene cell and nuclear entry properties of an HIV-1 based lentiviral vectors with the integration mechanism of the Sleeping Beauty transposon. The integrating activity of the HIV-1

integrase was prevented by the use of the catalytic mutant integrase D64V. It was found that non-integrating lentiviral vectors were able to deliver Sleeping Beauty transposon DNA and a functional transposase expression cassette to target cells, and that co-transduction with these components resulted in a 2-fold increase in vector integration relative to background. Chromosomal integration sites were recovered by ligation-mediated PCR and sequenced, demonstrating successful Sleeping Beauty transposition from a non-integrating lentiviral vector donor. Comparison of the integration pattern of the hybrid vector with that of a standard integrating lentiviral vector revealed a reduced frequency of integration in genes and no bias in integration towards actively transcribing genes when the hybrid vector was used.

The observed level of hybrid vector integration was relatively low, causing 2.5% of HeLa cells to become stably G418-resistant when transposase was provided compared to 1.3% background integration in the absence of transposase. Other authors using this Sleeping Beauty vector in plasmid transfections have reported efficiencies of 1-3% G418-resistant HeLa cells (585), so the efficiency of the hybrid vector is within the expected range. The high rate of background integration is most likely due to a combination of the high vector doses necessary to achieve detectable transposition together with the high background integration frequency of linear DNA such as non-integrating lentiviral vectors relative to circular DNA such as transfected plasmids. As calculated in Section 5.5, 40µl of transposon NILV and 5µl of transposase NILV were able to produce 1200 G418-resistant HeLa cells under optimised conditions and so have a combined integrating titre of 2.7×10^4 TU/ml. By contrast, a standard ILV with an identical backbone gave a G418-resistant HeLa titre of 2.9×10^8 TU/ml, a difference of four orders of magnitude. As with the hybrid system reported by Staunstrup *et al* (586), the efficiency of the system in this study might be improved by the use of the novel hyperactive transposase SB100X, possibly increasing the efficiency of transposition by 10- to 25-fold (587). However, the integration efficiency of the hybrid vector would most likely remain 2 to 3 orders of magnitude below that of standard integrating lentiviral vectors.

It is often difficult to compare integration efficiencies between authors because there are few standard assays or units of measurement. Integration efficiencies are commonly reported as ‘number of colonies’ or ‘x-fold increase in integration relative to background’ (588-593). The latter measure is likely to be particularly high in

plasmid transfection experiments as a result of the low rate of background integration of circular plasmid DNA in most cells. It is not clear why these measurements are often used in the literature rather than the more transparent '% of cells stably transduced', which is trivial to calculate from the same data.

Even the measure '% of cells stably transduced' can be misleading as it implies that 100% transduction is a perfect result. It is important to remember that this measure combines the efficiency of gene transfer to cells with the efficiency of integration. In experiments where an extremely high efficiency of gene transfer is readily obtained (e.g. plasmid transfection of 293T cells), a reasonably high '% of cells stably transduced' can be achieved even with an inefficient integrating system. A more accurate measure of integration efficiency would separate gene transfer and integration, for example by dividing the frequency of integration by the number of intracellular vector copies per cell after gene transfer. Although this is a grossly simple measure for a complex and heterogenous entity such as a cell, it could be used in combination with appropriate standards for cell type, transfection method, and means of quantifying integration efficiency.

Reporting efficiency results in relative terms can be confusing for those outside of the field as it leads to unrealistic expectations of what the currently available integrating systems are capable of in absolute terms. Indeed, the other integrating vector systems introduced in Section 2.3 appear to have significantly lower integration efficiencies than standard gammaretroviral and lentiviral vectors. Quantification of DNA intermediates during lentiviral vector transduction of cell lines showed that the total vector copy number peaked at 15-20 per cell while the integrated copy number peaked at 1-2, suggesting an efficiency of integration on the order of 10^{-1} integrations per intracellular vector genome (594). It has been estimated that the efficiency of both wildtype AAV site-specific integration by Rep and rAAV non-targeted integration is approximately 10^{-3} integrations per intracellular vector genome (595). In this study, the rate of Sleeping Beauty transposition due to the SB11 transposase was estimated to be 4×10^{-4} active chromosomal insertions per intracellular NILV genome, and the efficiency of the novel SB100X transposase is estimated to be 10- to 25-fold more efficient than this (596). The efficiency of insertion due to Flp recombinase was reported to be 2×10^{-5} per NILV 2-LTR circle (597). The efficiency of site-specific chromosomal insertion due to site-specific nucleases such as zinc finger nucleases has

not been reported in terms of insertions per intracellular genome, but '% transduction' values of 20-40% have been achieved in cell lines (598;599) and 0.06% of primary $CD34^+$ human haematopoietic stem cells (600). Although a side-by-side comparison has not yet been performed, conventional gammaretroviral vectors are able to transduce up to 60% of primary $CD34^+$ human haematopoietic stem cells (32).

In gene therapy applications where high copy number gene transfer is available, the relatively low efficiency of alternative integrating systems relative to retroviruses may not be rate limiting. For example, stable gene marking of 50% of human peripheral blood lymphocytes has been reported following electroporation with Sleeping Beauty plasmids (601). Interestingly, two recent reports using the novel transposase SB100X have shown Sleeping Beauty mediated gene marking of electroporated primary $CD34^+$ human haematopoietic stem cells which were subsequently able to reconstitute NOD-SCID γc-null mice (602;603). The authors reported 8% stable marking of human-derived cells in these animals (604) and the same integration sites were recovered from both lymphoid and myeloid lineages in individual mice, confirming transposition within a common progenitor (605). The risk of insertional mutagenesis due to this approach relative to gammaretroviral or lentiviral vectors will need to be confirmed in established genotoxicity assays (606;607).

In the final part of this study, the possibility of incorporating transposase protein into lentiviral vector particles in the same way as the HIV-1 integrase was investigated. It was found that heterologous proteins such as eGFP and the Sleeping Beauty transposase could be fused to the viral Gag polyprotein for packaging into vector particles, and that these polyproteins were processed by the viral protease as would be expected for Gag-Pol, the polyprotein which contains the HIV-1 integrase. However, no increase in the level of target cell integration could be detected when transposase was targeted into vector particles by fusion to the Gag protein, and this study did not demonstrate whether functional lentiviral particles containing transposase can be produced using this strategy. Assuming that the transposase can be targeted to virions and would be functional (though this must be shown directly), it is possible that transposition activity in target cells would require binding of transposase protein to the vector genome in sufficient quantities for transposition, just as the HIV integrase protein remains bound to the PIC/RTC throughout its transit to chromosomes. It is possible that continued association with the RTC/PIC after transduction could be

achieved by tethering transposase to a viral protein which is associated with the vector genome up to the point of integration, perhaps NC or even IN. The fact that retroviruses are able to package their own functional integrating protein, integrase, into virions shows that the approach has been made to work in nature. The ability to carrying integrating systems into cells as protein rather than nucleic acids may minimise any wider effects of the vector on the target cell, offering potential safety improvements in future integrating vectors.

At present, it seems likely that future generations of gene therapy vectors will incorporate heterologous mechanisms of transgene integration in order to improve the efficacy and/or safety of the gene therapy treatment. In this thesis it was shown that a hybrid lentivirus-Sleeping Beauty transposon vector can be generated which combines the cell entry properties of lentiviral vectors with the integration mechanism of Sleeping Beauty. As well as demonstrating that this class of hybrid vector is possible and functional, this study has highlighted a number of issues to be addressed in the future development of effective hybrid vectors. Significantly these include the requirement for efficient co-delivery of the integration mechanism to be employed and the importance of quantifying and optimising integration efficiency.

8. Appendix

8.1. Fusion Protein Amino Acid Sequences

8.1.1. Gag

MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQ
TGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAADTGHSNQVSQNYPIVQNI
QGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINE
EAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRM
YSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPGATLEE
MMTACQGVGGPGHKARVLAEAMSQVTNPATIMIQKGNFRNQRKTVKCFNCGKEGHIAKNCRAPRKKGCW
KCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPEESFRFGEETTTPSQKQEPIDKEL
YPLASLRSLFGSDPSSQ-

MA-CA-P2-NC

P1

P6

8.1.2. Gag-Pol

MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQ
TGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAADTGHSNQVSQNYPIVQNI
QGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINE
EAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRM
YSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPGATLEE
MMTACQGVGGPGHKARVLAEAMSQVTNPATIMIQKGNFRNQRKTVKCFNCGKEGHIAKNCRAPRKKGCW
KCGKEGHQMKDCTERQANFFREDLAFPQGKAREFSSEQTRANSPTRRELQVWGRDNNSLSEAGADRQGT
VSFSFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMNLPGRWKPKMIGGIGGFIKVGQYDQILI
EICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKAL
VEICTEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKQK
KSVTVLDVGDAYFSVPLDKDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQCSMTKILEPFR
KQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTV
QPIVLPEKDSWTVNDIQKLVGKLNWASQIYAGIKVRQLCKLLRGTKALTEVVPLTEEAELELAENREIL
KEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMKGAHTNDVKQLTEAVQKIATES
IVIWGKTPKFKLPIQKETWEAWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIIGAETFYVDGAANR
ETKLGKAGYVTDRGRQKVVPLTDTTNQKTELQAIHLALQDSGLEVNIVTDSQYALGIIQAQPDKSESEL
VSQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDGLVSAGIRKVLFLDGIDKAQEEHEKYHSNWRAMASDF
NLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQ

ETAYFLLKLAGRWPVKTVHTDNGSNFTSTTVKAACWWAGIKQEFGIPYNPQSQGVIESMNKELKKIIGQ
VRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRDP
VWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED-

MA-CA-P2-NC

PRO

POL

8.1.3. Gag-eGFP

MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQ
TGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAADTGHSNQVSQNYPIVQNI
QGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINE
EAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRM
YSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPGATLEE
MMTACQGVGGPGHKARVLAEAMSQVTNPATIMIQKGNFRNQRKTVKCFNCGKEGHIAKNCRAPRKKGCW
KCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPRGLQGMVSKGEELFTGVVPILVELDGDVNG
HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYV
QERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIK
VNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITL
GMDELYK-

MA-CA-P2-NC

P1

FIRST 5aa OF P6

LINKER

EGFP

8.1.4. Gag-SB11

MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQ
TGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAADTGHSNQVSQNYPIVQNI
QGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINE
EAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRM
YSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPGATLEE
MMTACQGVGGPGHKARVLAEAMSQVTNPATIMIQKGNFRNQRKTVKCFNCGKEGHIAKNCRAPRKKGCW
KCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPRGRSDIMGKSKEISQDLRKKIVDLHKSGSS

LGAISKRLKVPRSSVQTIVRKYKHHGTTQPSYRSGRRRVLSPRDERTLVRKVQINPRTTAKDLVKMLEE
TGTKVSISTVKRVLYRHNLKGRSARKKPLLQNRHKKARLRFARAHGDKDRTFWRNVLWSDETKIELFGH
NDHRYVWRKKGEACKPKNTIPTVKHGGGSIMLWGCFAAGGTGALHKIDGIMRKENYVDILKQHLKTSVR
KLKLGRKWVFQQDNDPKHTSKHVRKWLKDNKVKVLEWPSQSPDLNPIENLWAELKKRVRARRPTNLTQL
HQLCQEEWAKIHPTYCGKLVEGYPKRLTQVKQFKGNATKY-

MA-CA-P2-NC

P1

FIRST 5aa OF P6

LINKER

SB11 TRANSPOSASE

9. References

(1) Vink CA, Gaspar HB, Gabriel R, Schmidt M, McIvor RS, Thrasher AJ et al. Sleeping Beauty Transposition From Nonintegrating Lentivirus. Mol Ther 2009; 17(7):1197-1204.

(2) Steinman RM, Moberg CL. A triple tribute to the experiment that transformed biology. J Exp Med 1994; 179(2):379-384.

(3) Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol 1952; 36(1):39-56.

(4) Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953; 171(4356):737-738.

(5) Tatum EL. The determinants and evolution of life. PNAS 1964; 51:908-15.

(6) Gupta NK, Ohtsuka E, Weber H, Chang SH, Khorana HG. The joining of short deoxyribopolynucleotides by DNA-joining enzymes. PNAS 1968; 60(1):285-292.

(7) Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. PNAS 1973; 70(11):3240-3244.

(8) Friedmann T. A brief history of gene therapy. Nat Genet 1992; 2(2):93-98.

(9) Kay ERM. Incorporation of Deoxyribonucleic Acid by Mammalian Cells in vitro. Nature 1961; 191(4786):387-388.

(10) Graham FL, van der Eb AJ. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 1973; 52(2):456-467.

(11) Lawn RM, Fritsch EF, Parker RC, Blake G, Maniatis T. The isolation and characterization of linked delta- and beta-globin genes from a cloned library of human DNA. Cell 1978; 15(4):1157-1174.

(12) Green MR, Treisman R, Maniatis T. Transcriptional activation of cloned human beta-globin genes by viral immediate-early gene products. Cell 1983; 35(1):137-148.

(13) Cline MJ, Stang H, Mercola K, Morse L, Ruprecht R, Browne J et al. Gene transfer in intact animals. Nature 1980; 284(5755):422-425.

(14) Beutler E. The Cline Affair. Mol Ther 2001; 4(5):396-397.

(15) Hamer DH, Leder P. Expression of the chromosomal mouse beta-globin gene cloned in SV40. Nature 1979; 281(5726):35-40.

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

(17) Van Doren K, Hanahan D, Gluzman Y. Infection of eucaryotic cells by helper-independent recombinant adenoviruses: early region 1 is not obligatory for integration of viral DNA. J Virol 1984; 50(2):606-614.

(18) Hermonat PL, Muzyczka N. Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells. PNAS 1984; 81(20):6466-6470.

(19) Schaefer-Ridder M, Wang Y, Hofschneider PH. Liposomes as gene carriers: efficient transformation of mouse L cells by thymidine kinase gene. Science 1982; 215(4529):166-168.

(20) Neumann E, Schaefer-Ridder M, Wang Y, Hofschneider PH. Gene transfer into mouse lyoma cells by electroporation in high electric fields. EMBO J 1982; 1(7):841-845.

(21) Mann R, Mulligan RC, Baltimore D. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. Cell 1983; 33(1):153-159.

(22) Naldini L, Blomer U, Gallay P, Ory D, Mulligan R, Gage FH et al. In Vivo Gene Delivery and Stable Transduction of Nondividing Cells by a Lentiviral Vector. Science 1996; 272(5259):263-267.

(23) Boussif O, Lezoualc'h F, Zanta MA, Mergny MD, Scherman D, Demeneix B et al. A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. PNAS 1995; 92(16):7297-7301.

(24) Nathwani AC, Davidoff AM, Linch DC. A review of gene therapy for haematological disorders. Br J Haematol 2005; 128(1):3-17.

(25) Blaese RM, Culver KW, Miller AD, Carter CS, Fleisher T, Clerici M et al. T Lymphocyte-Directed Gene Therapy for ADA- SCID: Initial Trial Results After 4 Years. Science 1995; 270(5235):475-480.

(26) Antoine C, Muller S, Cant A, Cavazzana-Calvo M, Veys P, Vossen J et al. Long-term survival and transplantation of haemopoietic stem cells for immunodeficiencies: report of the European experience 1968-99. The Lancet 2003; 361(9357):553-560.

(27) Aiuti A, Vai S, Mortellaro A, Casorati G, Ficara F, Andolfi G et al. Immune reconstitution in ADA-SCID after PBL gene therapy and discontinuation of enzyme replacement. Nat Med 2002; 8(5):423-425.

(28) Raper SE, Chirmule N, Lee FS, Wivel NA, Bagg A, Gao Gp et al. Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. Molecular Genetics and Metabolism 2003; 80(1-2):148-158.

(29) Gene Therapy Clinical Trials Worldwide. J Gene Med. 2009.

(30) Cavazzana-Calvo M, Hacein-Bey S, Basile G, Gross F, Yvon E, Nusbaum P et al. Gene Therapy of Human Severe Combined Immunodeficiency (SCID)-X1 Disease. Science 2000; 288(5466):669-672.

(31) Gaspar HB, Parsley KL, Howe S, King D, Gilmour KC, Sinclair J et al. Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. The Lancet 2004; 364(9452):2181-2187.

(32) Hacein-Bey-Abina S, von Kalle C, Schmidt M, Le Deist F, Wulffraat N, McIntyre E et al. A Serious Adverse Event after Successful Gene Therapy for X-Linked Severe Combined Immunodeficiency. N Engl J Med 2003; 348(3):255-256.

(33) Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P et al. LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1. Science 2003; 302(5644):415-419.

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

(35) Alexander BL, Ali RR, Alton EW, Bainbridge JW, Braun S, Cheng SH, et al. Progress and Prospects: Gene Therapy Clinical Trials (Part 1). Gene Ther 2007; 14(20):1439-1447.

(36) Aiuti A, Bachoud-Lévi AC, Blesch A, Brenner MK, Cattaneo F, Chiocca EA et al. Progress and Prospects: Gene Therapy Clinical Trials (Part 2). Gene Ther 2007; 14(22):1555-1563.

(37) Aiuti A, Slavin S, Aker M, Ficara F, Deola S, Mortellaro A et al. Correction of ADA-SCID by Stem Cell Gene Therapy Combined with Nonmyeloablative Conditioning. Science 2002; 296(5577):2410-2413.

(38) Gaspar HB, Bjorkegren E, Parsley K, Gilmour KC, King D, Sinclair J et al. Successful Reconstitution of Immunity in ADA-SCID by Stem Cell Gene Therapy Following Cessation of PEG-ADA and Use of Mild Preconditioning. Mol Ther 2006; 14(4):505-513.

(39) Bonini C, Ferrari G, Verzeletti S, Servida P, Zappone E, Ruggieri L et al. HSV-TK Gene Transfer into Donor Lymphocytes for Control of Allogeneic Graft-Versus-Leukemia. Science 1997; 276(5319):1719-1724.

(40) Morgan RA, Dudley ME, Wunderlich JR, Hughes MS, Yang JC, Sherry RM et al. Cancer Regression in Patients After Transfer of Genetically Engineered Lymphocytes. Science 2006; 314(5796):126-129.

(41) Kaplitt MG, Feigin A, Tang C, Fitzsimons HL, Mattis P, Lawlor PA et al. Safety and tolerability of gene therapy with an adeno-associated virus (AAV) borne GAD gene for Parkinson's disease: an open label, phase I trial. The Lancet 2007; 369(9579):2097-2105.

(42) Bainbridge JWB, Smith AJ, Barker SS, Robbie S, Henderson R, Balaggan K et al. Effect of Gene Therapy on Visual Function in Leber's Congenital Amaurosis. N Engl J Med 2008; 358(21):2231-2239.

(43) Manno CS, Arruda VR, Pierce GF, Glader B, Ragni M, Rasko J et al. Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. Nat Med 2006; 12(3):342-347.

(44) Mitsuyasu RT, Merigan TC, Carr A, Zack JA, Winters MA, Workman C et al. Phase 2 gene therapy trial of an anti-HIV ribozyme in autologous CD34+ cells. Nat Med 2009; advanced online publication.

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

(46) Moreno-Carranza B, Gentsch M, Stein S, Schambach A, Santilli G, Rudolf E et al. Transgene optimization significantly improves SIN vector titers, gp91phox expression and reconstitution of superoxide production in X-CGD cells. Gene Ther 2008; 16(1):111-118.

(47) Griesenbach U, Alton EWFW. Gene transfer to the lung: Lessons learned from more than 2 decades of CF gene therapy. Advanced Drug Delivery Reviews 2009; In Press, Corrected Proof.

(48) Kohn DB, Candotti F. Gene Therapy Fulfilling Its Promise. N Engl J Med 2009; 360(5):518-521.

(49) Yee A, De Ravin SS, Elliott E, Ziegler JB. Severe combined immunodeficiency: a national surveillance study. Pediatr Allergy Immunol 2008; 19(4):298-302.

(50) Buckley RH. Molecular Defects in Human Severe Combined Immunodeficiency and Approaches to Immune Reconstitution. Annual Review of Immunology 2004; 22(1):625-655.

(51) Stephan V, Wahn V, Le Deist F, Dirksen U, Broker B, Muller-Fleckenstein I et al. Atypical X-Linked Severe Combined Immunodeficiency Due to Possible Spontaneous Reversion of the Genetic Defect in T Cells. N Engl J Med 1996; 335(21):1563-1567.

(52) Marshall E. Gene Therapy: Second Child in French Trial Is Found to Have Leukemia. Science 2003; 299(5605):320.

(53) Check E. Gene therapy put on hold as third child develops cancer. Nature 2005; 433(7026):561.

(54) Baum C. What Are the Consequences of the Fourth Case? Mol Ther 2007; 15(8):1401-1402.

(55) Lazo PA, Lee JS, Tsichlis PN. Long-Distance Activation of the Myc Protooncogene by Provirus Insertion in Mlvi-1 or Mlvi-4 in Rat T-Cell Lymphomas. PNAS 1990; 87(1):170-173.

(56) Wang CL, Wang BB, Bartha G+, Li L, Channa N, Klinger M et al. Activation of an oncogenic microRNA cistron by provirus integration. PNAS 2006; 103(49):18680-18684.

(57) Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. Nature 2005; 436(7048):272-276.

(58) Uren AG, Kool J, Matentzoglu K, de Ridder J, Mattison J, van Uitert M et al. Large-Scale Mutagenesis in p19ARF- and p53-Deficient Mice Identifies Cancer Genes and Their Collaborative Networks. Cell 2008; 133(4):727-741.

(59) Li Z, Dullmann J, Schiedlmeier B, Schmidt M, von Kalle C, Meyer J et al. Murine Leukemia Induced by Retroviral Gene Marking. Science 2002; 296(5567):497.

(60) Hacein-Bey-Abina S. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest 2008; 118(9):3132-3142.

(61) Pike-Overzet K, van der Burg M, Wagemaker G, van Dongen JJ, Staal FJ. New Insights and Unresolved Issues Regarding Insertional Mutagenesis in X-linked SCID Gene Therapy. Mol Ther 2007; 15(11):1910-1916.

(62) Guerra N, Tan YX, Joncker NT, Choy A, Gallardo F, Xiong N et al. NKG2D-Deficient Mice Are Defective in Tumor Surveillance in Models of Spontaneous Malignancy. Immunity 2008; 28(4):571-580.

(63) Baum C. Insertional mutagenesis in gene therapy and stem cell biology. Curr Opin Hematol 2007; 14(4):337-342.

(64) Seggewiss R, Pittaluga S, Adler RL, Guenaga FJ, Ferguson C, Pilz IH et al. Acute myeloid leukemia is associated with retroviral gene transfer to hematopoietic progenitor cells in a rhesus macaque. Blood 2006; 107(10):3865-3867.

(65) Pike-Overzet K, de Ridder D, Weerkamp F, Baert MRM, Verstegen MMA, Brugman MH et al. Ectopic retroviral expression of LMO2, but not IL2R-gamma, blocks human T-cell development from CD34+ cells: implications for leukemogenesis in gene therapy. Leukemia 2007; 21(4):754-763.

(66) Donsante A, Miller DG, Li Y, Vogler C, Brunt EM, Russell DW et al. AAV Vector Integration Sites in Mouse Hepatocellular Carcinoma. Science 2007; 317(5837):477.

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

(68) Zychlinski D, Schambach A, Modlich U, Maetzig T, Meyer J, Grassman E et al. Physiological Promoters Reduce the Genotoxic Risk of Integrating Gene Vectors. Mol Ther 2008; 16(4):718-725.

(69) Evans-Galea MV, Wielgosz MM, Hanawa H, Srivastava DK, Nienhuis AW. Suppression of Clonal Dominance in Cultured Human Lymphoid Cells by Addition of the cHS4 Insulator to a Lentiviral Vector. Mol Ther 2007; 15(4):801-809.

(70) Walisko O, Schorn A, Rolfs F, Devaraj A, Miskey C, Izsvak Z et al. Transcriptional Activities of the Sleeping Beauty Transposon and Shielding Its Genetic Cargo With Insulators. Mol Ther 2007; 16(2):359-369.

(71) Dave UP, Jenkins NA, Copeland NG. Gene Therapy Insertional Mutagenesis Insights. Science 2004; 303(5656):333.

(72) Woods NB, Bottero V, Schmidt M, von Kalle C, Verma IM. Gene therapy: Therapeutic gene causing lymphoma. Nature 2006; 440(7088):1123.

(73) Thrasher AJ, Gaspar HB, Baum C, Modlich U, Schambach A, Candotti F et al. Gene therapy: X-SCID transgene leukaemogenicity. Nature 2006; 443(7109):E5-E6.

(74) Wu X, Li Y, Crise B, Burgess SM. Transcription Start Regions in the Human Genome Are Favored Targets for MLV Integration. Science 2003; 300(5626):1749-1751.

(75) Schroder ARW, 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.

(76) Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, Kay MA. High-Resolution Genome-Wide Mapping of Transposon Integration in Mammals. Mol Cell Biol 2005; 25(6):2085-2094.

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

(78) Shiramizu B, Herndier BG, McGrath MS. Identification of a Common Clonal Human Immunodeficiency Virus Integration Site in Human Immunodeficiency Virus-associated Lymphomas. Cancer Res 1994; 54(8):2069-2072.

(79) Stephen SL, Sivanandam VG, Kochanek S. Homologous and heterologous recombination between adenovirus vector DNA and chromosomal DNA. J Gene Med 2008; 10(11):1176-1189.

(80) Coffin JM, Hughes SH, Varmus HE. Retroviruses. Plainview, New York: Cold Spring Harbor Laboratory Press, 1997.

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

(82) Zhou Q, Sharp PA. Novel mechanism and factor for regulation by HIV-1 Tat. EMBO J 1995; 14(2):321-328.

(83) Ratnasabapathy R, Sheldon M, Johal L, Hernandez N. The HIV-1 long terminal repeat contains an unusual element that induces the synthesis of short RNAs from various mRNA and snRNA promoters. Genes Dev 1990; 4(12A):2061-2074.

(84) Rana TM, Jeang KT. Biochemical and Functional Interactions between HIV-1 Tat Protein and TAR RNA. Archives of Biochemistry and Biophysics 1999; 365(2):175-185.

(85) Lochelt M, Flugel RM, Aboud M. The human foamy virus internal promoter directs the expression of the functional Bel 1 transactivator and Bet protein early after infection. J Virol 1994; 68(2):638-645.

(86) Furuichi Y, Shatkin AJ, Stavnezer E, Bishop JM. Blocked, methylated 5'-terminal sequence in avian sarcoma virus RNA. Nature 1975; 257(5527):618-620.

(87) Lai MM, Duesberg PH. Adenylic acid-rich sequence in RNAs of Rous sarcoma virus and Rauscher mouse leukaemia virus. Nature 1972; 235(5338):383-386.

(88) Purcell DF, Martin MA. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol 1993; 67(11):6365-6378.

(89) Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature 1989; 338(6212):254-257.

(90) Meyer BE, Meinkoth JL, Malim MH. Nuclear transport of human immunodeficiency virus type 1, visna virus, and equine infectious anemia virus Rev proteins: identification of a family of transferable nuclear export signals. J Virol 1996; 70(4):2350-2359.

(91) Smagulova F, Maurel S, Morichaud Z, Devaux C, Mougel M, Houzet L. The Highly Structured Encapsidation Signal of MuLV RNA is Involved in the Nuclear Export of its Unspliced RNA. J Mol Biol 2005; 354(5):1118-1128.

(92) D'Agostino DM, Felber BK, Harrison JE, Pavlakis GN. The Rev protein of human immunodeficiency virus type 1 promotes polysomal association and translation of gag/pol and vpu/env mRNAs. Mol Cell Biol 1992; 12(3):1375-1386.

(93) Kypr J, Mrazek J. Unusual codon usage of HIV. Nature 1987; 327(6117):20.

(94) Buck CB, Shen X, Egan MA, Pierson TC, Walker CM, Siliciano RF. The Human Immunodeficiency Virus Type 1 gag Gene Encodes an Internal Ribosome Entry Site. J Virol 2001; 75(1):181-191.

(95) Jamjoom GA, Naso RB, Arlinghaus RB. Further characterization of intracellular precursor polyproteins of Rauscher leukemia virus. Virology 1977; 78(1):11-34.

(96) Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature 1988; 331(6153):280-283.

(97) Jacks T. Translational suppression in gene expression in retroviruses and retrotransposons. Curr Top Microbiol Immunol 1990; 157:93-124.:93-124.

(98) Kollmus H, Honigman A, Panet A, Hauser H. The sequences of and distance between two cis-acting signals determine the efficiency of ribosomal frameshifting in human immunodeficiency virus type 1 and human T-cell leukemia virus type II in vivo. J Virol 1994; 68(9):6087-6091.

(99) Henderson LE, Krutzsch HC, Oroszlan S. Myristyl Amino-Terminal Acylation of Murine Retrovirus Proteins: An Unusual Post-Translational Protein Modification. PNAS 1983; 80(2):339-343.

(100) Zhou W, Parent LJ, Wills JW, Resh MD. Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. J Virol 1994; 68(4):2556-2569.

(101) Perez LG, Davis GL, Hunter E. Mutants of the Rous sarcoma virus envelope glycoprotein that lack the transmembrane anchor and cytoplasmic domains: analysis of intracellular transport and assembly into virions. J Virol 1987; 61(10):2981-2988.

(102) Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus

envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J Biol Chem 1990; 265(18):10373-10382.

(103) Li Y, Luo L, Rasool N, Kang CY. Glycosylation is necessary for the correct folding of human immunodeficiency virus gp120 in CD4 binding. J Virol 1993; 67(1):584-588.

(104) Pal R, Hoke GM, Sarngadharan MG. Role of Oligosaccharides in the Processing and Maturation of Envelope Glycoproteins of Human Immunodeficiency Virus Type 1. PNAS 1989; 86(9):3384-3388.

(105) Pantophlet R, Burton DR. GP120: Target for Neutralizing HIV-1 Antibodies. Annu Rev Immunol 2006; 24:739-69.:739-769.

(106) Zhu P, Chertova E, Bess J, Jr., Lifson JD, Arthur LO, Liu J et al. Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. PNAS 2003; 100(26):15812-15817.

(107) Gu M, Rappapor J, Leppla SH. Furin is important but not essential for the proteolytic maturation of gp160 of HIV-1. FEBS Letters 1995; 365(1):95-97.

(108) Gallaher WR. Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. Cell 1987; 50(3):327-328.

(109) Malim MH, Emerman M. HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell Host Microbe 2008; 3(6):388-398.

(110) Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 2002; 418(6898):646-650.

(111) Mariani R, Chen D, Schrofelbauer B, Navarro F, Konig R, Bollman B et al. Species-Specific Exclusion of APOBEC3G from HIV-1 Virions by Vif. Cell 2003; 114(1):21-31.

(112) Schrofelbauer B, Yu Q, Landau NR. New insights into the role of Vif in HIV-1 replication. AIDS Rev 2004; 6(1):34-39.

(113) Mehle A, Strack B, Ancuta P, Zhang C, McPike M, Gabuzda D. Vif Overcomes the Innate Antiviral Activity of APOBEC3G by Promoting Its Degradation in the Ubiquitin-Proteasome Pathway. J Biol Chem 2004; 279(9):7792-7798.

(114) Heinzinger NK, Bukrinsky MI, Haggerty SA, Ragland AM, Kewalramani V, Lee M et al. The Vpr Protein of Human Immunodeficiency Virus Type 1 Influences Nuclear Localization of Viral Nucleic Acids in Nondividing Host Cells. PNAS 1994; 91(15):7311-7315.

(115) Mansky LM, Preveral S, Selig L, Benarous R, Benichou S. The Interaction of Vpr with Uracil DNA Glycosylase Modulates the Human Immunodeficiency Virus Type 1 In Vivo Mutation Rate. J Virol 2000; 74(15):7039-7047.

(116) He J, Choe S, Walker R, Di Marzio P, Morgan DO, Landau NR. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. J Virol 1995; 69(11):6705-6711.

(117) Re F, Braaten D, Franke EK, Luban J. Human immunodeficiency virus type 1 Vpr arrests the cell cycle in G2 by inhibiting the activation of p34cdc2-cyclin B. J Virol 1995; 69(11):6859-6864.

(118) Felzien LK, Woffendin C, Hottiger MO, Subbramanian RA, Cohen EA, Nabel GJ. HIV transcriptional activation by the accessory protein, VPR, is mediated by the p300 co-activator. PNAS 1998; 95(9):5281-5286.

(119) Jacotot E, Ravagnan L, Loeffler M, Ferri KF, Vieira HLA, Zamzami N et al. The HIV-1 Viral Protein R Induces Apoptosis via a Direct Effect on the Mitochondrial Permeability Transition Pore. J Exp Med 2000; 191(1):33-46.

(120) Jacotot E, Ferri KF, El Hamel C, Brenner C, Druillennec S, Hoebeke J et al. Control of Mitochondrial Membrane Permeabilization by Adenine Nucleotide

Translocator Interacting with HIV-1 Viral Protein R and Bcl-2. J Exp Med 2001; 193(4):509-520.

(121) Bour S, Schubert U, Strebel K. The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: implications for the mechanism of degradation. J Virol 1995; 69(3):1510-1520.

(122) Margottin F, Bour SP, Durand H, Selig L, Benichou S, Richard V et al. A Novel Human WD Protein, h-beta-TrCP, that Interacts with HIV-1 Vpu Connects CD4 to the ER Degradation Pathway through an F-Box Motif. Molecular Cell 1998; 1(4):565-574.

(123) Neil SJD, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature 2008; 451(7177):425-430.

(124) Garcia JV, Miller AD. Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. Nature 1991; 350(6318):508-511.

(125) Schwartz O, Marechal V, Le Gall S, Lemonnier F, Heard JM. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. Nat Med 1996; 2(3):338-342.

(126) Stumptner-Cuvelette P, Morchoisne S, Dugast M, Le Gall S, Raposo G, Schwartz O et al. HIV-1 Nef impairs MHC class II antigen presentation and surface expression. PNAS 2001; 98(21):12144-12149.

(127) Pizzato M, Helander A, Popova E, Calistri A, Zamborlini A, Pal+¦ G et al. Dynamin 2 is required for the enhancement of HIV-1 infectivity by Nef. PNAS 2007; 104(16):6812-6817.

(128) Gheysen D, Jacobs E, de Foresta F, Thiriart C, Francotte M, Thines D et al. Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. Cell 1989; 59(1):103-112.

(129) Shioda T, Shibuta H. Production of human immunodeficiency virus (HIV)-like particles from cells infected with recombinant vaccinia viruses carrying the gag gene of HIV. Virology 1990; 175(1):139-148.

(130) Derdowski A, Ding L, Spearman P. A Novel Fluorescence Resonance Energy Transfer Assay Demonstrates that the Human Immunodeficiency Virus Type 1 Pr55Gag I Domain Mediates Gag-Gag Interactions. J Virol 2004; 78(3):1230-1242.

(131) Cen S, Niu M, Saadatmand J, Guo F, Huang Y, Nabel GJ et al. Incorporation of Pol into Human Immunodeficiency Virus Type 1 Gag Virus-Like Particles Occurs Independently of the Upstream Gag Domain in Gag-Pol. J Virol 2004; 78(2):1042-1049.

(132) Briggs JAG, Simon MN, Gross I, Krausslich HG, Fuller SD, Vogt VM et al. The stoichiometry of Gag protein in HIV-1. Nat Struct Mol Biol 2004; 11(7):672-675.

(133) Muller B, Tessmer U, Schubert U, Krausslich HG. Human Immunodeficiency Virus Type 1 Vpr Protein Is Incorporated into the Virion in Significantly Smaller Amounts than Gag and Is Phosphorylated in Infected Cells. J Virol 2000; 74(20):9727-9731.

(134) Paxton W, Connor RI, Landau NR. Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of gag and mutational analysis. J Virol 1993; 67(12):7229-7237.

(135) Pandori MW, Fitch NJ, Craig HM, Richman DD, Spina CA, Guatelli JC. Producer-cell modification of human immunodeficiency virus type 1: Nef is a virion protein. J Virol 1996; 70(7):4283-4290.

(136) Sova P, Volsky DJ, Wang L, Chao W. Vif Is Largely Absent from Human Immunodeficiency Virus Type 1 Mature Virions and Associates Mainly with Viral Particles Containing Unprocessed Gag. J Virol 2001; 75(12):5504-5517.

(137) Clever JL, Miranda D, Jr., Parslow TG. RNA Structure and Packaging Signals in the 5' Leader Region of the Human Immunodeficiency Virus Type 1 Genome. J Virol 2002; 76(23):12381-12387.

(138) Khorchid A, Javanbakht H, Wise S, Halwani R, Parniak MA, Wainberg MA et al. Sequences within Pr160gag-pol affecting the selective packaging of primer tRNALys3 into HIV-1. Journal of Molecular Biology 2000; 299(1):17-26.

(139) Ott DE. Potential roles of cellular proteins in HIV-1. Rev Med Virol 2002; 12(6):359-374.

(140) Zavada J. The pseudotypic paradox. J Gen Virol 1982; 63(Pt 1):15-24.

(141) Briggs JAG, Wilk T, Fuller SD. Do lipid rafts mediate virus assembly and pseudotyping? J Gen Virol 2003; 84(4):757-768.

(142) Gould SJ, Booth AM, Hildreth JEK. The Trojan exosome hypothesis. PNAS 2003; 100(19):10592-10597.

(143) Garrus JE, von Schwedler UK, Pornillos OW, Morham SG, Zavitz KH, Wang HE et al. Tsg101 and the Vacuolar Protein Sorting Pathway Are Essential for HIV-1 Budding. Cell 2001; 107(1):55-65.

(144) Raposo G, Moore M, Innes D, Leijendekker R, Leigh-Brown A, Benaroch P et al. Human macrophages accumulate HIV-1 particles in MHC II compartments. Traffic 2002; 3(10):718-729.

(145) Freed EO. HIV-1 Gag Proteins: Diverse Functions in the Virus Life Cycle. Virology 1998; 251(1):1-15.

(146) Pettit SC, Lindquist JN, Kaplan AH, Swanstrom R. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. Retrovirology 2005; 2:66.:66.

(147) Bukovsky AA, Dorfman T, Weimann A, Gottlinger HG. Nef association with human immunodeficiency virus type 1 virions and cleavage by the viral protease. J Virol 1997; 71(2):1013-1018.

(148) Cen S, Khorchid A, Gabor J, Rong L, Wainberg MA, Kleiman L. Roles of Pr55gag and NCp7 in tRNA3Lys Genomic Placement and the Initiation Step of Reverse Transcription in Human Immunodeficiency Virus Type 1. J Virol 2000; 74(22):10796-10800.

(149) Isel C, Marquet R, Keith G, Ehresmann C, Ehresmann B. Modified nucleotides of tRNA(3Lys) modulate primer/template loop-loop interaction in the

initiation complex of HIV-1 reverse transcription. J Biol Chem 1993; 268(34):25269-25272.

(150) Kleiman L. tRNA(Lys3): the primer tRNA for reverse transcription in HIV-1. IUBMB Life 2002; 53(2):107-114.

(151) Nisole S, Krust B, Callebaut C, Guichard G, Muller S, Briand JP et al. The Anti-HIV Pseudopeptide HB-19 Forms a Complex with the Cell-surface-expressed Nucleolin Independent of Heparan Sulfate Proteoglycans. J Biol Chem 1999; 274(39):27875-27884.

(152) Fortin JF, Cantin R, Tremblay MJ. T Cells Expressing Activated LFA-1 Are More Susceptible to Infection with Human Immunodeficiency Virus Type 1áParticles Bearing Host-Encoded ICAM-1. J Virol 1998; 72(3):2105-2112.

(153) Mondor I, Ugolini S, Sattentau QJ. Human Immunodeficiency Virus Type 1áAttachment to HeLa CD4 Cells Is CD4 Independent and gp120 Dependent and Requires Cell Surface Heparans. J Virol 1998; 72(5):3623-3634.

(154) Lin CL, Sewell AK, Gao GF, Whelan KT, Phillips RE, Austyn JM. Macrophage-tropic HIV Induces and Exploits Dendritic Cell Chemotaxis. J Exp Med 2000; 192(4):587-594.

(155) Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature 1984; 312(5996):763-767.

(156) Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M et al. Identification of a major co-receptor for primary isolates of HIV-1. Nature 1996; 381(6584):661-666.

(157) Doranz BJ, Rucker J, Yi Y, Smyth RJ, Samson M, Peiper SC et al. A Dual-Tropic Primary HIV-1 Isolate That Uses Fusin and the beta-Chemokine Receptors CKR-5, CKR-3, and CKR-2b as Fusion Cofactors. Cell 1996; 85(7):1149-1158.

(158) Choe H, Farzan M, Sun Y, Sullivan N, Rollins B, Ponath PD et al. The beta-Chemokine Receptors CCR3 and CCR5 Facilitate Infection by Primary HIV-1 Isolates. Cell 1996; 85(7):1135-1148.

(159) Alkhatib G, Combadiere C, Broder CC, Feng Y, Kennedy PE, Murphy PM et al. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science 1996; 272(5270):1955-1958.

(160) Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA et al. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature 1996; 381(6584):667-673.

(161) Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science 1996; 272(5263):872-877.

(162) Labrijn AF, Poignard P, Raja A, Zwick MB, Delgado K, Franti M et al. Access of Antibody Molecules to the Conserved Coreceptor Binding Site on Glycoprotein gp120 Is Sterically Restricted on Primary Human Immunodeficiency Virus Type 1. J Virol 2003; 77(19):10557-10565.

(163) Bosch ML, Earl PL, Fargnoli K, Picciafuoco S, Giombini F, Wong-Staal F et al. Identification of the fusion peptide of primate immunodeficiency viruses. Science 1989; 244(4905):694-697.

(164) Carr CM, Kim PS. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. Cell 1993; 73(4):823-832.

(165) Colman PM, Lawrence MC. The structural biology of Type I viral membrane fusion. Nat Rev Mol Cell Biol 2003; 4(4):309-319.

(166) Markosyan RM, Cohen FS, Melikyan GB. HIV-1 Envelope Proteins Complete Their Folding into Six-helix Bundles Immediately after Fusion Pore Formation. Mol Biol Cell 2003; 14(3):926-938.

(167) Marechal V, Prevost MC, Petit C, Perret E, Heard JM, Schwartz O. Human Immunodeficiency Virus Type 1 Entry into Macrophages Mediated by Macropinocytosis. J Virol 2001; 75(22):11166-11177.

(168) Fredericksen BL, Wei BL, Yao J, Luo T, Garcia JV. Inhibition of Endosomal/Lysosomal Degradation Increases the Infectivity of Human Immunodeficiency Virus. J Virol 2002; 76(22):11440-11446.

(169) Daecke J, Fackler OT, Dittmar MT, Krausslich HG. Involvement of Clathrin-Mediated Endocytosis in Human Immunodeficiency Virus Type 1 Entry. J Virol 2005; 79(3):1581-1594.

(170) Hovanessian AG, Briand JP, Said EA, Svab J, Ferris S, Dali H et al. The Caveolin-1 Binding Domain of HIV-1 Glycoprotein gp41 Is an Efficient B Cell Epitope Vaccine Candidate against Virus Infection. Immunity 2004; 21(5):617-627.

(171) Hug P, Lin HMJ, Korte T, Xiao X, Dimitrov DS, Wang JM et al. Glycosphingolipids Promote Entry of a Broad Range of Human Immunodeficiency Virus Type 1 Isolates into Cell Lines Expressing CD4, CXCR4, and/or CCR5. J Virol 2000; 74(14):6377-6385.

(172) Manes S, del Real G, Lacalle RA, Lucas P, Gomez-Mouton C, Sanchez-Palomino S et al. Membrane raft microdomains mediate lateral assemblies required for HIV-1 infection. EMBO Rep 2000; 1(2):190-196.

(173) Popik W, Alce TM, Au WC. Human Immunodeficiency Virus Type 1 Uses Lipid Raft-Colocalized CD4 and Chemokine Receptors for Productive Entry into CD4+ T Cells. J Virol 2002; 76(10):4709-4722.

(174) Percherancier Y, Lagane B, Planchenault T, Staropoli I, Altmeyer R, Virelizier JL et al. HIV-1 Entry into T-cells Is Not Dependent on CD4 and CCR5 Localization to Sphingolipid-enriched, Detergent-resistant, Raft Membrane Domains. J Biol Chem 2003; 278(5):3153-3161.

(175) Schaeffer E, Geleziunas R, Greene WC. Human Immunodeficiency Virus Type 1 Nef Functions at the Level of Virus Entry by Enhancing Cytoplasmic Delivery of Virions. J Virol 2001; 75(6):2993-3000.

(176) Zheng YH, Plemenitas A, Fielding CJ, Peterlin BM. Nef increases the synthesis of and transports cholesterol to lipid rafts and HIV-1 progeny virions. PNAS 2003; 100(14):8460-8465.

(177) Tobiume M, Lineberger JE, Lundquist CA, Miller MD, Aiken C. Nef Does Not Affect the Efficiency of Human Immunodeficiency Virus Type 1 Fusion with Target Cells. J Virol 2003; 77(19):10645-10650.

(178) Cavrois M, Neidleman J, Yonemoto W, Fenard D, Greene WC. HIV-1 virion fusion assay: uncoating not required and no effect of Nef on fusion. Virology 2004; 328(1):36-44.

(179) Chazal N, Singer G, Aiken C, Hammarskjold ML, Rekosh D. Human Immunodeficiency Virus Type 1 Particles Pseudotyped with Envelope Proteins That Fuse at Low pH No Longer Require Nef for Optimal Infectivity. J Virol 2001; 75(8):4014-4018.

(180) Bukrinskaya A, Brichacek B, Mann A, Stevenson M. Establishment of a Functional Human Immunodeficiency Virus Type 1 (HIV-1) Reverse Transcription Complex Involves the Cytoskeleton. J Exp Med 1998; 188(11):2113-2125.

(181) Kaushik R, Ratner L. Role of Human Immunodeficiency Virus Type 1 Matrix Phosphorylation in an Early Postentry Step of Virus Replication. J Virol 2004; 78(5):2319-2326.

(182) Miller MD, Farnet CM, Bushman FD. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. J Virol 1997; 71(7):5382-5390.

(183) Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, Stevenson M. Association of Integrase, Matrix, and Reverse Transcriptase Antigens of Human Immunodeficiency Virus Type 1 with Viral Nucleic Acids Following Acute Infection. PNAS 1993; 90(13):6125-6129.

(184) Fassati A, Goff SP. Characterization of Intracellular Reverse Transcription Complexes of Human Immunodeficiency Virus Type 1. J Virol 2001; 75(8):3626-3635.

(185) Aiken C, Trono D. Nef stimulates human immunodeficiency virus type 1 proviral DNA synthesis. J Virol 1995; 69(8):5048-5056.

(186) Campbell EM, Nunez R, Hope TJ. Disruption of the Actin Cytoskeleton Can Complement the Ability of Nef To Enhance Human Immunodeficiency Virus Type 1 Infectivity. J Virol 2004; 78(11):5745-5755.

(187) Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. The cytoplasmic body component TRIM5-alpha restricts HIV-1 infection in Old World monkeys. Nature 2004; 427(6977):848-853.

(188) Towers G. The control of viral infection by tripartite motif proteins and cyclophilin A. Retrovirology 2007; 4(1):40.

(189) de Soultrait VR, Caumont A, Durrens P, Calmels C, Parissi V, Recordon P et al. HIV-1 integrase interacts with yeast microtubule-associated proteins. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 2002; 1575(1-3):40-48.

(190) Wilk T, Gowen B, Fuller SD. Actin Associates with the Nucleocapsid Domain of the Human Immunodeficiency Virus Gag Polyprotein. J Virol 1999; 73(3):1931-1940.

(191) McDonald D, Vodicka MA, Lucero G, Svitkina TM, Borisy GG, Emerman M et al. Visualization of the intracellular behavior of HIV in living cells. J Cell Biol 2002; 159(3):441-452.

(192) Iordanskiy S, Berro R, Altieri M, Kashanchi F, Bukrinsky M. Intracytoplasmic maturation of the human immunodeficiency virus type 1 reverse transcription complexes determines their capacity to integrate into chromatin. Retrovirology 2006; 3:4.:4.

(193) Yu SF, Sullivan MD, Linial ML. Evidence that the Human Foamy Virus Genome Is DNA. J Virol 1999; 73(2):1565-1572.

(194) DeStefano JJ, Buiser RG, Mallaber LM, Bambara RA, Fay PJ. Human immunodeficiency virus reverse transcriptase displays a partially processive 3' to 5' endonuclease activity. J Biol Chem 1991; 266(36):24295-24301.

(195) Kati WM, Johnson KA, Jerva LF, Anderson KS. Mechanism and fidelity of HIV reverse transcriptase. J Biol Chem 1992; 267(36):25988-25997.

(196) DeStefano JJ, Mallaber LM, Fay PJ, Bambara RA. Quantitative analysis of RNA cleavage during RNA-directed DNA synthesis by human immunodeficiency and avian myeloblastosis virus reverse transcriptases. Nucleic Acids Res 1994; 22(18):3793-3800.

(197) Laakso MM, Sutton RE. Replicative fidelity of lentiviral vectors produced by transient transfection. Virology In Press, Corrected Proof.

(198) Ratner L, Haseltine W, Patarca R, Livak KJ, Starcich B, Josephs SF et al. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature 1985; 313(6000):277-284.

(199) Lanchy JM, Keith G, Le Grice SF, Ehresmann B, Ehresmann C, Marquet R. Contacts between Reverse Transcriptase and the Primer Strand Govern the Transition from Initiation to Elongation of HIV-1 Reverse Transcription. J Biol Chem 1998; 273(38):24425-24432.

(200) Ohi Y, Clever JL. Sequences in the 5' and 3' R Elements of Human Immunodeficiency Virus Type 1 Critical for Efficient Reverse Transcription. J Virol 2000; 74(18):8324-8334.

(201) Varmus HE, Heasley S, Kung HJ, Oppermann H, Smith VC, Bishop JM et al. Kinetics of synthesis, structure and purification of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells. Journal of Molecular Biology 1978; 120(1):55-82.

(202) van Wamel JLB, Berkhout B. The First Strand Transfer during HIV-1 Reverse Transcription Can Occur either Intramolecularly or Intermolecularly. Virology 1998; 244(2):245-251.

(203) Allain B, Rascle JB, de Rocquigny H, Roques B, Darlix JL. CIS elements and Trans-acting factors required for minus strand DNA transfer during reverse transcription of the genomic RNA of murine leukemia virus. Journal of Molecular Biology 1998; 277(2):225-235.

(204) Topping R, Demoitie MA, Shin NH, Telesnitsky A. Cis-acting elements required for strong stop acceptor template selection during moloney murine leukemia virus reverse transcription. Journal of Molecular Biology 1998; 281(1):1-15.

(205) Berkhout B, Vastenhouw NL, Klasens BI, Huthoff H. Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. RNA 2001; 7(8):1097-1114.

(206) Chen Y, Balakrishnan M, Roques BP, Bambara RA. Steps of the Acceptor Invasion Mechanism for HIV-1 Minus Strand Strong Stop Transfer. J Biol Chem 2003; 278(40):38368-38375.

(207) Rausch JW, Le Grice SF. 'Binding, bending and bonding': polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus. Int J Biochem Cell Biol 2004; 36(9):1752-1766.

(208) Gotte M, Maier G, Onori AM, Cellai L, Wainberg MA, Heumann H. Temporal Coordination between Initiation of HIV (+)-Strand DNA Synthesis and Primer Removal. J Biol Chem 1999; 274(16):11159-11169.

(209) Pullen KA, Ishimoto LK, Champoux JJ. Incomplete removal of the RNA primer for minus-strand DNA synthesis by human immunodeficiency virus type 1 reverse transcriptase. J Virol 1992; 66(1):367-373.

(210) Wakefield JK, Morrow CD. Mutations within the Primer Binding Site of the Human Immunodeficiency Virus Type 1 Define Sequence Requirements Essential for Reverse Transcription. Virology 1996; 220(2):290-298.

(211) Charneau P, Clavel F. A single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. J Virol 1991; 65(5):2415-2421.

(212) Charneau P, Mirambeau G, Roux P, Paulous S, Buc H, Clavel F. HIV-1 Reverse Transcription A Termination Step at the Center of the Genome. Journal of Molecular Biology 1994; 241(5):651-662.

(213) Lewis PF, Emerman M. Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. J Virol 1994; 68(1):510-516.

(214) Roe T, Reynolds TC, Yu G, Brown PO. Integration of murine leukemia virus DNA depends on mitosis. EMBO J 1993; 12(5):2099-2108.

(215) Bukrinsky M. A hard way to the nucleus. Mol Med 2004; 10(1-6):1-5.

(216) Sirven A, Pflumio F, Zennou V, Titeux M, Vainchenker W, Coulombel L et al. The human immunodeficiency virus type-1 central DNA flap is a crucial determinant for lentiviral vector nuclear import and gene transduction of human hematopoietic stem cells. Blood 2000; 96(13):4103-4110.

(217) Dvorin JD, Bell P, Maul GG, Yamashita M, Emerman M, Malim MH. Reassessment of the Roles of Integrase and the Central DNA Flap in Human Immunodeficiency Virus Type 1 Nuclear Import. J Virol 2002; 76(23):12087-12096.

(218) Limon A, Nakajima N, Lu R, Ghory HZ, Engelman A. Wild-Type Levels of Nuclear Localization and Human Immunodeficiency Virus Type 1 Replication in the Absence of the Central DNA Flap. J Virol 2002; 76(23):12078-12086.

(219) Bukrinsky MI, Haggerty S, Dempsey MP, Sharova N, Adzhubei A, Spitz L et al. A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. Nature 1993; 365(6447):666-669.

(220) Reil H, Bukovsky AA, Gelderblom HR, Gottlinger HG. Efficient HIV-1 replication can occur in the absence of the viral matrix protein. EMBO J 1998; 17(9):2699-2708.

(221) Bouyac-Bertoia M, Dvorin JD, Fouchier RAM, Jenkins Y, Meyer BE, Wu LI et al. HIV-1 Infection Requires a Functional Integrase NLS. Molecular Cell 2001; 7(5):1025-1035.

(222) Limon A, Devroe E, Lu R, Ghory HZ, Silver PA, Engelman A. Nuclear Localization of Human Immunodeficiency Virus Type 1 Preintegration Complexes (PICs): V165A and R166A Are Pleiotropic Integrase Mutants Primarily Defective for Integration, Not PIC Nuclear Import. J Virol 2002; 76(21):10598-10607.

(223) Yamashita M, Emerman M. The Cell Cycle Independence of HIV Infections Is Not Determined by Known Karyophilic Viral Elements. PLoS Pathog 2005; 1(3):e18.

(224) Yamashita M, Emerman M. Capsid Is a Dominant Determinant of Retrovirus Infectivity in Nondividing Cells. J Virol 2004; 78(11):5670-5678.

(225) Hindmarsh P, Leis J. Retroviral DNA Integration. Microbiol Mol Biol Rev 1999; 63(4):836-843.

(226) Engelman A, Craigie R. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. J Virol 1992; 66(11):6361-6369.

(227) Gao K, Wong S, Bushman F. 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 2004; 78(13):6715-6722.

(228) Engelman A, Hickman AB, Craigie R. The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. J Virol 1994; 68(9):5911-5917.

(229) Appa RS, Shin CG, Lee P, Chow SA. Role of the Nonspecific DNA-binding Region and alpha Helices within the Core Domain of Retroviral Integrase in Selecting Target DNA Sites for Integration. J Biol Chem 2001; 276(49):45848-45855.

(230) Harper AL, Sudol M, Katzman M. An Amino Acid in the Central Catalytic Domain of Three Retroviral Integrases That Affects Target Site Selection in Nonviral DNA. J Virol 2003; 77(6):3838-3845.

(231) Brown PO, Bowerman B, Varmus HE, Bishop JM. Correct integration of retroviral DNA in vitro. Cell 1987; 49(3):347-356.

(232) Lobel LI, Murphy JE, Goff SP. The palindromic LTR-LTR junction of Moloney murine leukemia virus is not an efficient substrate for proviral integration. J Virol 1989; 63(6):2629-2637.

(233) Van Maele B, Busschots K, Vandekerckhove L, Christ F, Debyser Z. Cellular co-factors of HIV-1 integration. Trends in Biochemical Sciences 2006; 31(2):98-105.

(234) Fujiwara T, Mizuuchi K. Retroviral DNA integration: Structure of an integration intermediate. Cell 1988; 54(4):497-504.

(235) Craigie R, Fujiwara T, Bushman F. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell 1990; 62(4):829-837.

(236) Yoder KE, Bushman FD. Repair of Gaps in Retroviral DNA Integration Intermediates. J Virol 2000; 74(23):11191-11200.

(237) Engelman A, Cherepanov P. The Lentiviral Integrase Binding Protein LEDGF/p75 and HIV-1 Replication. PLoS Pathog 2008; 4(3):e1000046.

(238) 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. J Biol Chem 2004; 279(47):48883-48892.

(239) Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH et al. An Essential Role for LEDGF/p75 in HIV Integration. Science 2006; 314(5798):461-464.

(240) Wu X, Li Y, Crise B, Burgess SM, Munroe DJ. Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. J Virol 2005; 79(8):5211-5214.

(241) Holman AG, Coffin JM. Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. PNAS 2005; 102(17):6103-6107.

(242) Correction for Holman et al., Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites, PNAS 2005 102:6103-6107. PNAS 2005; 102(17):6238.

(243) Bushman FD. Tethering Human Immunodeficiency Virus 1 Integrase to a DNA Site Directs Integration to Nearby Sequences. PNAS 1994; 91(20):9233-9237.

(244) Pryciak PM, Varmus HE. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell 1992; 69(5):769-780.

(245) 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 Res 2007; 17(8):1186-1194.

(246) Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, Hannenhalli S et al. Genome-wide Analysis of Retroviral DNA Integration. Nat Rev Micro 2005; 3(11):848-858.

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

(248) Trobridge GD, Miller DG, Jacobs MA, Allen JM, Kiem HP, Kaul R et al. Foamy virus vector integration sites in normal human cells. PNAS 2006; 103(5):1498-1503.

(249) Trobridge G, Josephson N, Vassilopoulos G, Mac J, Russell DW. Improved foamy virus vectors with minimal viral sequences. Mol Ther 2002; 6(3):321-328.

(250) Donahue RE, Kessler SW, Bodine D, McDonagh K, Dunbar C, Goodman S et al. Helper virus induced T cell lymphoma in nonhuman primates after retroviral mediated gene transfer. J Exp Med 1992; 176(4):1125-1135.

(251) Zufferey R, Donello JE, Trono D, Hope TJ. Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element Enhances Expression of Transgenes Delivered by Retroviral Vectors. J Virol 1999; 73(4):2886-2892.

(252) Kingsman SM, Mitrophanous K, Olsen JC. Potential oncogene activity of the woodchuck hepatitis post-transcriptional regulatory element (WPRE). 2004; 12(1):3-4.

(253) Schambach A, Bohne J, Baum C, Hermann FG, Egerer L, von Laer D et al. Woodchuck hepatitis virus post-transcriptional regulatory element deleted from X protein and promoter sequences enhances retroviral vector titer and expression. Gene Ther 2005; 13(7):641-645.

(254) Yu SF, von Roden T, Kantoff PW, Garber C, Seiberg M, Rother U et al. Self-inactivating retroviral vectors designed for transfer of whole genes into mammalian cells. PNAS 1986; 83(10):3194-3198.

(255) Zufferey R, Dull T, Mandel RJ, Bukovsky A, Quiroz D, Naldini L et al. Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery. J Virol 1998; 72(12):9873-9880.

(256) Goverdhana S, Puntel M, Xiong W, Zirger JM, Barcia C, Curtin JF et al. Regulatable Gene Expression Systems for Gene Therapy Applications: Progress and Future Challenges. Mol Ther 2005; 12(2):189-211.

(257) Brown BD, Gentner B, Cantore A, Colleoni S, Amendola M, Zingale A et al. Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. Nat Biotech 2007; 25(12):1457-1467.

(258) Zufferey R, Nagy D, Mandel RJ, Naldini L, Trono D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. Nat Biotech 1997; 15(9):871-875.

(259) Dull T, Zufferey R, Kelly M, Mandel RJ, Nguyen M, Trono D et al. A Third-Generation Lentivirus Vector with a Conditional Packaging System. J Virol 1998; 72(11):8463-8471.

(260) Kotsopoulou E, Kim VN, Kingsman AJ, Kingsman SM, Mitrophanous KA. A Rev-Independent Human Immunodeficiency Virus Type 1 (HIV-1)-Based Vector That Exploits a Codon-Optimized HIV-1 gag-pol Gene. J Virol 2000; 74(10):4839-4852.

(261) Leavitt AD, Robles G, Alesandro N, Varmus HE. Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection. J Virol 1996; 70(2):721-728.

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

(263) Burns JC, Friedmann T, Driever W, Burrascano M, Yee J. Vesicular Stomatitis Virus G Glycoprotein Pseudotyped Retroviral Vectors: Concentration to Very High Titer and Efficient Gene Transfer into Mammalian and Nonmammalian Cells. PNAS 1993; 90(17):8033-8037.

(264) Pichlmair A, Diebold SS, Gschmeissner S, Takeuchi Y, Ikeda Y, Collins MK et al. Tubulovesicular Structures within Vesicular Stomatitis Virus G Protein-Pseudotyped Lentiviral Vector Preparations Carry DNA and Stimulate Antiviral Responses via Toll-Like Receptor 9. J Virol 2007; 81(2):539-547.

(265) Schlegel R, Tralka TS, Willingham MC, Pastan I. Inhibition of VSV binding and infectivity by phosphatidylserine: Is phosphatidylserine a VSV-binding site? Cell 1983; 32(2):639-646.

(266) Coil DA, Miller AD. Phosphatidylserine Is Not the Cell Surface Receptor for Vesicular Stomatitis Virus. J Virol 2004; 78(20):10920-10926.

(267) Lewinski MK, Bisgrove D, Shinn P, Chen H, Hoffmann C, Hannenhalli S et al. Genome-Wide Analysis of Chromosomal Features Repressing Human Immunodeficiency Virus Transcription. J Virol 2005; 79(11):6610-6619.

(268) Bayer M, Kantor B, Cockrell A, Ma H, Zeithaml B, Li X et al. A Large U3 Deletion Causes Increased In Vivo Expression From a Nonintegrating Lentiviral Vector. Mol Ther 2008.

(269) Cornu TI, Cathomen T. Targeted Genome Modifications Using Integrase-deficient Lentiviral Vectors. Mol Ther 2007; 15(12):2107-2113.

(270) Emery DW, Yannaki E, Tubb J, Stamatoyannopoulos G. A chromatin insulator protects retrovirus vectors from chromosomal position effects. PNAS 2000; 97(16):9150-9155.

(271) Ellis J, Yao S. Retrovirus silencing and vector design: relevance to normal and cancer stem cells? Curr Gene Ther 2005; 5(4):367-373.

(272) Gilbert JR, Wong-Staal F. HIV-2 and SIV Vector Systems. Somatic Cell and Molecular Genetics 2001; 26(1):83-98.

(273) Sauter SL, Gasmi M. FIV Vector Systems. Somatic Cell and Molecular Genetics 2001; 26(1):99-129.

(274) Olsen JC. EIAV, CAEV and Other Lentivirus Vector Systems. Somatic Cell and Molecular Genetics 2001; 26(1 - 6):131-145.

(275) Kotin RM, Siniscalco M, Samulski RJ, Zhu XD, Hunter L, Laughlin CA et al. Site-specific integration by adeno-associated virus. PNAS 1990; 87(6):2211-2215.

(276) Im DS, Muzyczka N. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell 1990; 61(3):447-457.

(277) Weitzman MD, Kyostio SR, Kotin RM, Owens RA. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. PNAS 1994; 91(13):5808-5812.

(278) McCarty DM, Young SM, Samulski RJ. Integration of adeno-associated virus (AAV) and recombinant AAV vectors. Ann Rev Gen 2004; 38(1):819-845.

(279) Vincent-Lacaze N, Snyder RO, Gluzman R, Bohl D, Lagarde C, Danos O. Structure of Adeno-Associated Virus Vector DNA following Transduction of the Skeletal Muscle. J Virol 1999; 73(3):1949-1955.

(280) Nakai H, Iwaki Y, Kay MA, Couto LB. Isolation of Recombinant Adeno-Associated Virus Vector-Cellular DNA Junctions from Mouse Liver. J Virol 1999; 73(7):5438-5447.

(281) Miller DG, Trobridge GD, Petek LM, Jacobs MA, Kaul R, Russell DW. Large-Scale Analysis of Adeno-Associated Virus Vector Integration Sites in Normal Human Cells. J Virol 2005; 79(17):11434-11442.

(282) Ivics Z, Hackett PB, Plasterk RH, Izsvak Z. Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. Cell 1997; 91(4):501-510.

(283) Izsvak Z, Khare D, Behlke J, Heinemann U, Plasterk RH, Ivics Z. Involvement of a Bifunctional, Paired-like DNA-binding Domain and a Transpositional Enhancer in Sleeping Beauty Transposition. J Biol Chem 2002; 277(37):34581-34588.

(284) Moldt B, Yant SR, Andersen PR, Kay MA, Mikkelsen JG. Cis-Acting Gene Regulatory Activities in the Terminal Regions of Sleeping Beauty DNA Transposon-Based Vectors. Human Gene Therapy 2007; 18(12):1193-1204.

(285) Ivics Z, Izsvak Z, Minter A, Hackett PB. Identification of functional domains and evolution of Tc1-like transposable elements. PNAS 1996; 93(10):5008-5013.

(286) Plasterk RHA, Izsvak Z, Ivics Z. Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends in Genetics 1999; 15(8):326-332.

(287) Cui Z, Geurts AM, Liu G, Kaufman CD, Hackett PB. Structure-Function Analysis of the Inverted Terminal Repeats of the Sleeping Beauty Transposon. Journal of Molecular Biology 2002; 318(5):1221-1235.

(288) Ikeda R, Kokubu C, Yusa K, Keng VW, Horie K, Takeda J. Sleeping Beauty Transposase Has an Affinity for Heterochromatin Conformation. Mol Cell Biol 2007; 27(5):1665-1676.

(289) Yusa K, Takeda J, Horie K. Enhancement of Sleeping Beauty Transposition by CpG Methylation: Possible Role of Heterochromatin Formation. Mol Cell Biol 2004; 24(9):4004-4018.

(290) Vos JC, De B, I, Plasterk RH. Transposase is the only nematode protein required for in vitro transposition of Tc1. Genes Dev 1996; 10(6):755-761.

(291) Yant SR, Ehrhardt A, Mikkelsen JG, Meuse L, Pham T, Kay MA. Transposition from a gutless adeno-transposon vector stabilizes transgene expression in vivo. Nat Biotech 2002; 20(10):999-1005.

(292) Zayed H, Izsvak Z, Khare D, Heinemann U, Ivics Z. The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. Nucl Acids Res 2003; 31(9):2313-2322.

(293) Luo G, Ivics Z, Izsvak Z, Bradley A. Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. PNAS 1998; 95(18):10769-10773.

(294) Yant SR, Kay MA. Nonhomologous-End-Joining Factors Regulate DNA Repair Fidelity during Sleeping Beauty Element Transposition in Mammalian Cells. Mol Cell Biol 2003; 23(23):8505-8518.

(295) Izsvak Z, Stuwe EE, Fiedler D, Katzer A, Jeggo PA, Ivics Z. Healing the Wounds Inflicted by Sleeping Beauty Transposition by Double-Strand Break Repair in Mammalian Somatic Cells. Molecular Cell 2004; 13(2):279-290.

(296) Walisko O, Izsvak Z, Szabo K, Kaufman CD, Herold S, Ivics Z. Sleeping Beauty transposase modulates cell-cycle progression through interaction with Miz-1. PNAS 2006; 103(11):4062-4067.

(297) Liu G, Aronovich EL, Cui Z, Whitley CB, Hackett PB. Excision of Sleeping Beauty transposons: parameters and applications to gene therapy. J Gene Med 2004; 6(5):574-583.

(298) Geurts AM, Hackett CS, Bell JB, Bergemann TL, Collier LS, Carlson CM et al. Structure-based prediction of insertion-site preferences of transposons into chromosomes. Nucl Acids Res 2006; 34(9):2803-2811.

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

(300) Mikkelsen JG, Yant SR, Meuse L, Huang Z, Xu H, Kay MA. Helper-Independent Sleeping Beauty Transposon-Transposase Vectors for Efficient Nonviral Gene Delivery and Persistent Gene Expression in Vivo. Mol Ther 2003; 8(4):654-665.

(301) Wilber A, Frandsen JL, Geurts JL, Largaespada DA, Hackett PB, McIvor RS. RNA as a Source of Transposase for Sleeping Beauty-Mediated Gene Insertion and Expression in Somatic Cells and Tissues. Mol Ther 2006; 13(3):625-630.

(302) Geurts AM, Yang Y, Clark KJ, Liu G, Cui Z, Dupuy AJ et al. Gene transfer into genomes of human cells by the sleeping beauty transposon system. Molecular Therapy 2003; 8(1):108-117.

(303) Lohe AR, Hartl DL. Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. Mol Biol Evol 1996; 13(4):549-555.

(304) Izsvak Z, Ivics Z, Plasterk RH. Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. Journal of Molecular Biology 2000; 302(1):93-102.

(305) Baus J, Liu L, Heggestad AD, Sanz S, Fletcher BS. Hyperactive Transposase Mutants of the Sleeping Beauty Transposon. Molecular Therapy 2005; 12(6):1148-1156.

(306) Yant SR, Park J, Huang Y, Mikkelsen JG, Kay MA. Mutational Analysis of the N-Terminal DNA-Binding Domain of Sleeping Beauty Transposase: Critical Residues for DNA Binding and Hyperactivity in Mammalian Cells. Mol Cell Biol 2004; 24(20):9239-9247.

(307) Zayed H, Izsvak Z, Walisko O, Ivics Z. Development of Hyperactive Sleeping Beauty Transposon Vectors by Mutational Analysis. Molecular Therapy 2004; 9(2):292-304.

(308) Liu L, Sanz S, Heggestad AD, Antharam V, Notterpek L, Fletcher BS. Endothelial Targeting of the Sleeping Beauty Transposon within Lung. Mol Ther 2004; 10(1):97-105.

(309) Mates L, Chuah MKL, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A et al. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat Genet 2009; 41(6):753-761.

(310) Garrison BS, Yant SR, Mikkelsen JG, Kay MA. Postintegrative Gene Silencing within the Sleeping Beauty Transposition System. Mol Cell Biol 2007; 27(24):8824-8833.

(311) Dalsgaard T, Moldt B, Sharma N, Wolf G, Schmitz A, Pedersen FS et al. Shielding of Sleeping Beauty DNA Transposon-delivered Transgene Cassettes by Heterologous Insulators in Early Embryonal Cells. Mol Ther 2008; 17(1):121-130.

(312) Sijen T, Plasterk RHA. Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. Nature 2003; 426(6964):310-314.

(313) Yant SR, Meuse L, Chiu W, Ivics Z, Izsvak Z, Kay MA. Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nat Genet 2000; 25(1):35-41.

(314) Ohlfest JR, Frandsen JL, Fritz S, Lobitz PD, Perkinson SG, Clark KJ et al. Phenotypic correction and long-term expression of factor VIII in hemophilic mice by immunotolerization and nonviral gene transfer using the Sleeping Beauty transposon system. Blood 2005; 105(7):2691-2698.

(315) Ohlfest JR, Lobitz PD, Perkinson SG, Largaespada DA. Integration and Long-Term Expression in Xenografted Human Glioblastoma Cells Using a Plasmid-Based Transposon System. Mol Ther 2004; 10(2):260-268.

(316) Williams DA. Sleeping Beauty Vector System Moves Toward Human Trials in the United States. Mol Ther 2008; 16(9):1515-1516.

(317) Singh H, Manuri PR, Olivares S, Dara N, Dawson MJ, Huls H et al. Redirecting Specificity of T-Cell Populations For CD19 Using the Sleeping Beauty System. Cancer Res 2008; 68(8):2961-2971.

(318) Carlson CM, Frandsen JL, Kirchhof N, McIvor RS, Largaespada DA. Somatic integration of an oncogene-harboring Sleeping Beauty transposon models liver tumor development in the mouse. PNAS 2005; 102(47):17059-17064.

(319) Miskey C, Izsvak Z, Plasterk RH, Ivics Z. The Frog Prince: a reconstructed transposon from Rana pipiens with high transpositional activity in vertebrate cells. Nucl Acids Res 2003; 31(23):6873-6881.

(320) Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T. Efficient Transposition of the piggyBac (PB) Transposon in Mammalian Cells and Mice. Cell 2005; 122(3):473-483.

(321) Kawakami K, Noda T. Transposition of the Tol2 Element, an Ac-Like Element From the Japanese Medaka Fish Oryzias latipes, in Mouse Embryonic Stem Cells. Genetics 2004; 166(2):895-899.

(322) Wilson MH, Coates CJ, George AL. PiggyBac Transposon-mediated Gene Transfer in Human Cells. Mol Ther 2007; 15(1):139-145.

(323) Sternberg N, Sauer B, Hoess R, Abremski K. Bacteriophage P1 cre gene and its regulatory region: Evidence for multiple promoters and for regulation by DNA methylation. Journal of Molecular Biology 1986; 187(2):197-212.

(324) Hartley JL, Donelson JE. Nucleotide sequence of the yeast plasmid. Nature 1980; 286(5776):860-864.

(325) Rausch H, Lehmann M. Structural analysis of the bacteriophage phiC31 attachment site. Nucl Acids Res 1991; 19(19):5187-5189.

(326) Kuhstoss S, Rao RN. Analysis of the integration function of the streptomycete bacteriophage phiC31. Journal of Molecular Biology 1991; 222(4):897-908.

(327) Grindley NDF, Whiteson KL, Rice PA. Mechanisms of Site-Specific Recombination. Ann Rev Biochem 2006; 75(1):567-605.

(328) Smith MC, Thorpe HM. Diversity in the serine recombinases. Mol Microbiol 2002; 44(2):299-307.

(329) Groth AC, Olivares EC, Thyagarajan B, Calos MP. A phage integrase directs efficient site-specific integration in human cells. PNAS 2000; 97(11):5995-6000.

(330) Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B, Kirby PJ, Hillman RT et al. Integration Specificity of Phage phiC31 Integrase in the Human Genome. Journal of Molecular Biology 2006; 357(1):28-48.

(331) Ehrhardt A, Engler JA, Xu H, Cherry AM, Kay MA. Molecular Analysis of Chromosomal Rearrangements in Mammalian Cells After phiC31-Mediated Integration. Human Gene Therapy 2006; 17(11):1077-1094.

(332) Liu J, Jeppesen I, Nielsen K, Jensen TG. phiC31 integrase induces chromosomal aberrations in primary human fibroblasts. Gene Ther 2006; 13(15):1188-1190.

(333) Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP. Site-Specific Genomic Integration in Mammalian Cells Mediated by Phage phiC31 Integrase. Mol Cell Biol 2001; 21(12):3926-3934.

(334) Keravala A, Lee S, Thyagarajan B, Olivares EC, Gabrovsky VE, Woodard LE et al. Mutational Derivatives of phiC31 Integrase With Increased Efficiency and Specificity. Mol Ther 2008; 17(1):112-120.

(335) Doetschman T, Gregg RG, Maeda N, Hooper ML, Melton DW, Thompson S et al. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature 1987; 330(6148):576-578.

(336) Capecchi MR. Generating mice with targeted mutations. Nat Med 2001; 7(10):1086-1090.

(337) Kostriken R, Strathern JN, Klar AJS, Hicks JB, Heffron F. A site-specific endonuclease essential for mating-type switching in Saccharomyces cerevisiae. Cell 1983; 35(1):167-174.

(338) Rouet P, Smih F, Jasin M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. PNAS 1994; 91(13):6064-6068.

(339) Sung P, Klein H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Nat Rev Mol Cell Biol 2006; 7(10):739-750.

(340) Paques F, Duchateau P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr Gene Ther 2007; 7(1):49-66.

(341) Pavletich NP, Pabo CO. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 1991; 252(5007):809-817.

(342) Segal DJ, Dreier B, Beerli RR, Barbas CF. Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5GC-GNN-3GC DNA target sequences. PNAS 1999; 96(6):2758-2763.

(343) Dreier B, Fuller RP, Segal DJ, Lund CV, Blancafort P, Huber A et al. Development of Zinc Finger Domains for Recognition of the 5'-CNN-3' Family DNA Sequences and Their Use in the Construction of Artificial Transcription Factors. J Biol Chem 2005; 280(42):35588-35597.

(344) Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M et al. Rapid "Open-Source" Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification. Molecular Cell 2008; 31(2):294-301.

(345) Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. PNAS 1996; 93(3):1156-1160.

(346) Smith J, Bibikova M, Whitby FG, Reddy AR, Chandrasegaran S, Carroll D. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucl Acids Res 2000; 28(17):3361-3369.

(347) Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 2005; 435(7042):646-651.

(348) Moehle EA, Rock JM, Lee YL, Jouvenot Y, DeKelver RC, Gregory PD et al. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. PNAS 2007; 104(9):3055-3060.

(349) Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA et al. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotech 2007; 25(11):1298-1306.

(350) Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, Liu O et al. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotech 2008; 26(7):808-816.

(351) Alwin S, Gere MB, Guhl E, Effertz K, Barbas CF, Segal DJ et al. Custom Zinc-Finger Nucleases for Use in Human Cells. Mol Ther 2005; 12(4):610-617.

(352) Porteus MH. Mammalian Gene Targeting with Designed Zinc Finger Nucleases. Mol Ther 2006; 13(2):438-446.

(353) Gouble A, Smith J, Bruneau S, Perez C, Guyot V, Cabaniols JP et al. Efficient in toto targeted recombination in mouse liver by meganuclease-induced double-strand break. J Gene Med 2006; 8(5):616-622.

(354) Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ, Cathomen T. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotech 2007; 25(7):786-793.

(355) Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I et al. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotech 2007; 25(7):778-785.

(356) Lambowitz AM, Zimmerly S. Mobile Group II Introns. Ann Rev Gen 2004; 38(1):1-35.

(357) Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D et al. Retrohoming of a Bacterial Group II Intron: Mobility via Complete Reverse Splicing, Independent of Homologous DNA Recombination. Cell 1998; 94(4):451-462.

(358) Matsuura M, Saldanha R, Ma H, Wank H, Yang J, Mohr G et al. A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within theGÇëintron. Genes & Development 1997; 11(21):2910-2924.

(359) Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM. Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nat Biotech 2001; 19(12):1162-1167.

(360) Perutka J, Wang W, Goerlitz D, Lambowitz AM. Use of Computer-designed Group II Introns to Disrupt Escherichia coli DExH/D-box Protein and DNA Helicase Genes. Journal of Molecular Biology 2004; 336(2):421-439.

(361) Guo H, Karberg M, Long M, Jones JP, III, Sullenger B, Lambowitz AM. Group II Introns Designed to Insert into Therapeutically Relevant DNA Target Sites in Human Cells. Science 2000; 289(5478):452-457.

(362) Mastroianni M, Watanabe K, White TB, Zhuang F, Vernon J, Matsuura M et al. Group II Intron-Based Gene Targeting Reactions in Eukaryotes. PLoS ONE 2008; 3(9):e3121.

(363) 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.

(364) Tan W, Dong Z, Wilkinson TA, Barbas CF, III, 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. J Virol 2006; 80(4):1939-1948.

(365) Wilson MH, Kaminski JM, George J. Functional zinc finger/sleeping beauty transposase chimeras exhibit attenuated overproduction inhibition. FEBS Letters 2005; 579(27):6205-6209.

(366) Yant SR, Huang Y, Akache B, Kay MA. Site-directed transposon integration in human cells. Nucl Acids Res 2007; 35(7):e50.

(367) Ivics Z, Katzer A, Stuwe EE, Fiedler D, Knespel S, Izsvak Z. Targeted Sleeping Beauty Transposition in Human Cells. Mol Ther 2007; 15(6):1137-1144.

(368) Staunstrup NH, Moldt B, Mates L, Villesen P, Jakobsen M, Ivics Z et al. Hybrid Lentivirus-transposon Vectors With a Random Integration Profile in Human Cells. Mol Ther 2009, Epub ahead of print.

(369) Bowers WJ, Mastrangelo MA, Howard DF, Southerland HA, Maguire-Zeiss KA, Federoff HJ. Neuronal Precursor-Restricted Transduction via in Utero CNS Gene Delivery of a Novel Bipartite HSV Amplicon/Transposase Hybrid Vector. Mol Ther 2006; 13(3):580-588.

(370) Murphy SJ, Chong H, Bell S, Diaz RM, Vile RG. Novel Integrating Adenoviral/Retroviral Hybrid Vector for Gene Therapy. Human Gene Therapy 2002; 13(6):745-760.

(371) Shojitanaka A, Mizuochi T, Komuro K. Gene Transfer Using Purified Retroviral Integrase. Biochemical and Biophysical Research Communications 1994; 203(3):1756-1764.

(372) Heister T, Heid I, Ackermann M, Fraefel C. Herpes Simplex Virus Type 1/Adeno-Associated Virus Hybrid Vectors Mediate Site-Specific Integration at the Adeno-Associated Virus Preintegration Site, AAVS1, on Human Chromosome 19. J Virol 2002; 76(14):7163-7173.

(373) Wang Y, Camp SM, Niwano M, Shen X, Bakowska JC, Breakefield XO et al. Herpes Simplex Virus Type 1/Adeno-Associated Virus rep+ Hybrid Amplicon Vector Improves the Stability of Transgene Expression in Human Cells by Site-Specific Integration. J Virol 2002; 76(14):7150-7162.

(374) Recchia A, Parks RJ, Lamartina S, Toniatti C, Pieroni L, Palombo F et al. Site-specific integration mediated by a hybrid adenovirus/adeno-associated virus vector. PNAS 1999; 96(6):2615-2620.

(375) Zhang C, Cortez NG, Berns KI. Characterization of a Bipartite Recombinant Adeno-Associated Viral Vector for Site-Specific Integration. Human Gene Therapy 2007; 18(9):787-797.

(376) Surosky RT, Urabe M, Godwin SG, McQuiston SA, Kurtzman GJ, Ozawa K et al. Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. J Virol 1997; 71(10):7951-7959.

(377) Moldt B, Staunstrup NH, Jakobsen M, Yanez-Munoz RJ, Mikkelsen JG. Genomic insertion of lentiviral DNA circles directed by the yeast Flp recombinase. BMC Biotechnol 2008; 8(1):60.

(378) Sauer B, Henderson N. Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. New Biol 1990; 2(5):441-449.

(379) Koch KS, Aoki T, Wang Y, Atkinson AE, Gleiberman AS, Glebov OK et al. Site-specific integration of targeted DNA into animal cell genomes. Gene 2000; 249(1-2):135-144.

(380) Ehrhardt A, Yant SR, Giering JC, Xu H, Engler JA, Kay MA. Somatic Integration From an Adenoviral Hybrid Vector into a Hot Spot in Mouse Liver Results in Persistent Transgene Expression Levels In Vivo. Mol Ther 2007; 15(1):146-156.

(381) Cornu TI, Cathomen T. Targeted Genome Modifications Using Integrase-deficient Lentiviral Vectors. Mol Ther 2007; 15(12):2107-2113.

(382) Miller DG, Petek LM, Russell DW. Human Gene Targeting by Adeno-Associated Virus Vectors Is Enhanced by DNA Double-Strand Breaks. Mol Cell Biol 2003; 23(10):3550-3557.

(383) Porteus MH, Cathomen T, Weitzman MD, Baltimore D. Efficient Gene Targeting Mediated by Adeno-Associated Virus and DNA Double-Strand Breaks. Mol Cell Biol 2003; 23(10):3558-3565.

(384) Cornu TI, Cathomen T. Targeted Genome Modifications Using Integrase-deficient Lentiviral Vectors. Mol Ther 2007; 15(12):2107-2113.

(385) Santoni de Sio FR, Cascio P, Zingale A, Gasparini M, Naldini L. Proteasome activity restricts lentiviral gene transfer into hematopoietic stem cells and is down-regulated by cytokines that enhance transduction. Blood 2006; 107(11):4257-4265.

(386) Nightingale SJ, Hollis RP, Pepper KA, Petersen D, Yu XJ, Yang C et al. Transient Gene Expression by Nonintegrating Lentiviral Vectors. Mol Ther 2006; 13(6):1121-1132.

(387) Ketting RF, Fischer SEJ, Plasterk RH. Target choice determinants of the Tc1 transposon of Caenorhabditis elegans. Nucl Acids Res 1997; 25(20):4041-4047.

(388) Hollis RP, Nightingale SJ, Wang X, Pepper KA, Yu XJ, Barsky L et al. Stable gene transfer to human CD34(+) hematopoietic cells using the Sleeping Beauty transposon. Exp Hematol 2006; 34(10):1333-1343.

(389) Demaison C, Parsley K, Brouns G, Scherr M, Battmer K, Kinnon C et al. High-Level Transduction and Gene Expression in Hematopoietic Repopulating Cells Using a Human Imunodeficiency Virus Type 1-Based Lentiviral Vector Containing an Internal Spleen Focus Forming Virus Promoter. Human Gene Therapy 2002; 13(7):803-813.

(390) Inoue H, Nojima H, Okayama H. High efficiency transformation of Escherichia coli with plasmids. Gene 1990; 96(1):23-28.

(391) Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucl Acids Res 1988; 16(3):1215.

(392) Wu X, Burgess SM. Integration target site selection for retroviruses and transposable elements. Cell Mol Life Sci 2004; 61(19-20):2588-2596.

(393) Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005; 437(7057):376-380.

(394) Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM et al. The Human Genome Browser at UCSC. Genome Res 2002; 12(6):996-1006.

(395) Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A Sequence Logo Generator. Genome Res 2004; 14(6):1188-1190.

(396) Lamprecht MR, Sabatini DM, Carpenter AE. CellProfiler: free, versatile software for automated biological image analysis. Biotechniques 2007; 42(1):71-75.

(397) Yanez-Munoz RJ, Balaggan KS, MacNeil A, Howe SJ, Schmidt M, Smith AJ et al. Effective gene therapy with nonintegrating lentiviral vectors. Nat Med 2006; 12(3):348-353.

(398) Kumar R, Vandegraaff N, Mundy L, Burrell CJ, Li P. Evaluation of PCR-based methods for the quantitation of integrated HIV-1 DNA. Journal of Virological Methods 2002; 105(2):233-246.

(399) Apolonia L, Waddington SN, Fernandes C, Ward NJ, Bouma G, Blundell MP et al. Stable Gene Transfer to Muscle Using Non-integrating Lentiviral Vectors. Mol Ther 2007; 15(11):1947-1954.

(400) Geraerts M, Willems S, Baekelandt V, Debyser Z, Gijsbers R. Comparison of lentiviral vector titration methods. BMC Biotechnology 2006; 6(1):34.

(401) Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green Fluorescent Protein as a Marker for Gene Expression. Science 1994; 263(5148):802-805.

(402) DuBridge RB, Tang P, Hsia HC, Leong PM, Miller JH, Calos MP. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. Mol Cell Biol 1987; 7(1):379-387.

(403) Macville M, Schrock E, Padilla-Nash H, Keck C, Ghadimi BM, Zimonjic D et al. Comprehensive and Definitive Molecular Cytogenetic Characterization of HeLa Cells by Spectral Karyotyping. Cancer Res 1999; 59(1):141-150.

(404) Brinster RL, Chen HY, Trumbauer ME, Yagle MK, Palmiter RD. Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. PNAS 1985; 82(13):4438-4442.

(405) Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. PNAS 1977; 74(12):5463-5467.

(406) Ellison V, Abrams H, Roe T, Lifson J, Brown P. Human immunodeficiency virus integration in a cell-free system. J Virol 1990; 64(6):2711-2715.

(407) Shannon CE. A mathematical theory of communication. Bell System Tech J 1948; 27:379-423.

(408) Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M et al. The UCSC Genome Browser Database: 2008 update. Nucl Acids Res 2008; 36(S1):D773-D779.

(409) 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 Biol 2004; 2(8):E234.

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

(411) Philippe S, Sarkis C, Barkats M, Mammeri H, Ladroue C, Petit C et al. Lentiviral vectors with a defective integrase allow efficient and sustained transgene expression in vitro and in vivo. PNAS 2006; 103(47):17684-17689.

(412) Butler SL, Hansen MST, Bushman FD. A quantitative assay for HIV DNA integration in vivo. Nat Med 2001; 7(5):631-634.

(413) Fawell S, Seery J, Daikh Y, Moore C, Chen LL, Pepinsky B et al. Tat-mediated delivery of heterologous proteins into cells. PNAS 1994; 91(2):664-668.

(414) Mann DA, Frankel AD. Endocytosis and targeting of exogenous HIV-1 Tat protein. EMBO J 1991; 10(7):1733-1739.

(415) Vives E, Brodin P, Lebleu B. A Truncated HIV-1 Tat Protein Basic Domain Rapidly Translocates through the Plasma Membrane and Accumulates in the Cell Nucleus. J Biol Chem 1997; 272(25):16010-16017.

(416) Derossi D, Joliot AH, Chassaing G, Prochiantz A. The third helix of the Antennapedia homeodomain translocates through biological membranes. J Biol Chem 1994; 269(14):10444-10450.

(417) Rojas M, Donahue JP, Tan Z, Lin YZ. Genetic engineering of proteins with cell membrane permeability. Nat Biotech 1998; 16(4):370-375.

(418) Wadia JS, Stan RV, Dowdy SF. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. Nat Med 2004; 10(3):310-315.

(419) Gump JM, Dowdy SF. TAT transduction: the molecular mechanism and therapeutic prospects. Trends Mol Med 2007; 13(10):443-448.

(420) Lamartina S, Roscilli G, Rinaudo D, Delmastro P, Toniatti C. Lipofection of Purified Adeno-Associated Virus Rep68 Protein: toward a Chromosome-Targeting Nonviral Particle. J Virol 1998; 72(9):7653-7658.

(421) Xu Y, Liu S, Yu G, Chen J, Chen J, Xu X et al. Excision of selectable genes from transgenic goat cells by a protein transducible TAT-Cre recombinase. Gene 2008; 419(1-2):70-74.

(422) Hubner W, Chen P, Portillo AD, Liu Y, Gordon RE, Chen BK. Sequence of Human Immunodeficiency Virus Type 1 (HIV-1) Gag Localization and Oligomerization Monitored with Live Confocal Imaging of a Replication-Competent, Fluorescently Tagged HIV-1. J Virol 2007; 81(22):12596-12607.

(423) Muller B, Daecke J, Fackler OT, Dittmar MT, Zentgraf H, Krausslich HG. Construction and Characterization of a Fluorescently Labeled Infectious Human Immunodeficiency Virus Type 1 Derivative. J Virol 2004; 78(19):10803-10813.

(424) Munk C, Brandt SM, Lucero G, Landau NR. A dominant block to HIV-1 replication at reverse transcription in simian cells. PNAS 2002; 99(21):13843-13848.

(425) Link N, Aubel C, Kelm JM, Marty RR, Greber D, Djonov V et al. Therapeutic protein transduction of mammalian cells and mice by nucleic acid-free lentiviral nanoparticles. Nucl Acids Res 2006; 34(2):e16.

(426) Peretti S, Schiavoni I, Pugliese K, Federico M. Cell Death Induced by the Herpes Simplex Virus-1 Thymidine Kinase Delivered by Human Immunodeficiency Virus-1-Based Virus-like Particles. Mol Ther 2005; 12(6):1185-1196.

(427) Sheng N, Pettit SC, Tritch RJ, Ozturk DH, Rayner MM, Swanstrom R et al. Determinants of the human immunodeficiency virus type 1 p15NC-RNA interaction that affect enhanced cleavage by the viral protease. J Virol 1997; 71(8):5723-5732.

(428) Craven RC, Leure-duPree AE, Erdie CR, Wilson CB, Wills JW. Necessity of the spacer peptide between CA and NC in the Rous sarcoma virus gag protein. J Virol 1993; 67(10):6246-6252.

(429) Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, Klein CA et al. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. J Virol 1994; 68(12):8017-8027.

(430) Wiegers K, Rutter G, Kottler H, Tessmer U, Hohenberg H, Krausslich HG. Sequential Steps in Human Immunodeficiency Virus Particle Maturation Revealed by Alterations of Individual Gag Polyprotein Cleavage Sites. J Virol 1998; 72(4):2846-2854.

(431) Izsvak Z, Ivics Z. Sleeping Beauty Transposition: Biology and Applications for Molecular Therapy. Molecular Therapy 2004; 9(2):147-156.

(432) Kumarasuriyar A, Dombrowski C, Rider DA, Nurcombe V, Cool SM. A novel use of TAT-EGFP to validate techniques to alter osteosarcoma cell surface glycosaminoglycan expression. J Mol Histol 2007; 38(5):435-447.

(433) Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, Kay MA. High-resolution genome-wide mapping of transposon integration in mammals. Mol Cell Biol 2005; 25(6):2085-2094.

(434) Peng PD, Cohen CJ, Yang S, Hsu C, Jones S, Zhao Y et al. Efficient nonviral Sleeping Beauty transposon-based TCR gene transfer to peripheral blood lymphocytes confers antigen-specific antitumor reactivity. Gene Ther 2009, Epub ahead of print.

(435) Xue X, Huang X, Nodland SE, Mates L, Ma L, Izsvak Z et al. Stable gene transfer and expression in cord blood-derived CD34+ hematopoietic stem and

progenitor cells by a hyperactive Sleeping Beauty transposon system. Blood 2009, Epub ahead of print.