INVESTIGATION OF THE EFFECTS OF VIRUS INTEGRATION ON HOST GENE EXPRESSION IN MOUSE TUMOUR SAMPLES

A thesis submitted for the degree of Doctor of Philosophy

by
Emma Awele Osejindu

School of Health and Social Care
Brunel University
October 2011

ABSTRACT

Clonally derived liver tumours and an ovarian cyst developed in mice following EIAV and FIV delivery *in utero*. LAM PCR and 454 sequencing was used to retrieve proviral insertion sites. TaqMan analysis revealed gene expression changes in lentiviral infected tumours. STRING and IPA networks identified links between genes flanking the lentivirus provirus and oncogenic pathways supporting the role of insertional mutagenesis in Hepatocellular Carcinoma (HCC).

Global methylation analysis demonstrated increased relative methylation levels in lentivirus (EIAV, FIV, and HIV) infected normal and tumour samples. This provided strong evidence for host defence against lentivirus infection by epigenetic means. Microarray data showed altered expression of *Dnmt1* and *Dnmt3b* and TaqMan analysis revealed specific changes in *Dnmts* levels when compared to uninfected liver.

The evidence found for involvement of DNA methylation associated with lentivirus infection and possibly tumour development required that this study be repeated *in vitro*. DNA methylation was investigated at early time points after lentivirus and retrovirus infection in HepG2 cells. Results revealed sharp increases in global methylation and *Dnmt* levels at 24 and 30hrs post infection. E2F targets play a key role in the regulation of gene expression and aberrations result in the development of cancer.

Of the 94 E2F target genes analysed 77.7% were involved in DNA damage and repair mechanisms, 21.3% were known oncogenes or shown to exert oncogenic activity and 80.9% were categorised as HCC target genes.

The fact that all lentiviral/retroviral vectors used in this study were found to cause changes in methylation and gene expression *in vivo* and *in vitro* suggests that these vectors, at least in the mouse, are genotoxic.

Findings here support the use of the fetal animal model to identify vector genotoxicity and the mechanisms of lentiviral vector-induced tumorigenesis.

This model may be a valuable tool to evaluate the safety of lentiviruses for gene therapy.

TABLE OF CONTENTS

Page number

ABSTRACT……………………………………………………….ii

TABLE OF CONTENTS……………………………………....iii

LIST OF FIGURES……………………………………………....iv

LIST OF TABLES……………………………………………...v

ABBREVIATIONS……………………………………………vi

ACKNOWLEDGEMENTS…………………………….………vii

DECLARATION…………………………………………….......viii

CHAPTER 1

1 INTRODUCTION……………………….……………………..1-40

1.1 Gene therapy………………………….……………………………...1

1.1.1 Retrovirus: an overview……………………….……………….........2-6

1.1.2 The retrovirus life cycle………………………………….................7-11

1.1.3 Lentiviruses……………………………………………………..…….11-13

1.1.3.1 Human immunodeficiency virus (HIV) ……………….………13-14

1.1.3.2 Equine infectious anemia virus (EIAV)…………….…….................14

1.1.3.3 Feline immunodeficiency virus (FIV) ………………..…….….........15

1.1.4 Retroviruses and lentiviruses in gene therapy……………………......16-17

1.1.4.1 The production of retroviral/lentiviral vectors for gene therapy...17-19

1.1.4.2 Successful gene therapy clinical trials using viral vectors a history……………………………………………………………..19-21

1.1.5 Retroviruses and genotoxicity in the host……………….....…………**21-23**

1.1.5.1 Genotoxicity following retrovirus and lentivirus infection of the host genome……………...…………………………….…..…..**23**

1.1.5.2 Retroviruses and aberrant splicing……………………………........**24**

1.1.5.3 Retroviruses and promoter activation………………………………**24**

1.1.5.4 Retroviruses and enhancer activation………………………………**25**

1.1.5.5 Retroviruses and read through transcription………………..…...**25-26**

1.1.6 Retrovirus genotoxicity in gene therapy trials……….......................**26-31**

1.1.6.1 X – Linked SCID trials…………………………………….......**26-29**

1.1.6.2 Chronic Granulomatous Disease (CGD) trial………….........**30-31**

1.1.7 Models for retrovirus genotoxicity……………………………........**32-35**

1.1.7.1 Fetal gene therapy model……………………………………....**35-37**

1.1.8 Modifications to the design of retroviral vectors to avoid insertional mutagenesis………………………….................**38-39**

1.1.9 Hypothesis……………………………………………………………**39-40**

CHAPTER 2

MATERIALS AND METHODS……………………….............41-77

2.1 MATERIALS…………………………………………………41-51

2.1.1 General chemicals………………………………………….………41

2.1.2 Transformation reagents……………………………......................41

2.1.3 Ligation amplification mediated (LAM PCR) reagents……………......42

2.1.4 Oligonucleotides used for LAM PCR…………………………………..43

2.1.5 Reagents for RNA extraction and for RT PCR and QPCR analysis…....44

2.1.6 Reagents for *DNase I* treatment……………………………………...44

2.1.7 CDNA synthesis reagents for RT (Reverse Transciptase) PCR analysis…………………………………………...…………....44

2.1.8 RT PCR reagents…………………………………………………......44

2.1.9 Oligonucleotides for RT- PCR analysis……………………………...45

2.1.10 RNA purification for Real - Time Quantitative (Q) – PCR reagents…..45

2.1.11 CDNA synthesis for Real - Time Quantitative (Q) – PCR reagents…....45

2.1.12 QPCR reagents for gene expression analysis and vector copy number (VCN)…………………………………………….....46

2.1.13 QPCR primer probe sets for VCN analysis……………..…....……….....46

2.1.14 Pre designed and custom made TaqMan probes for gene expression analysis – Applied Biosystems…………………………...47

2.1.15 Tissue culture reagents…………………………………………….....48

2.1.16 Reagents for DNA extraction from frozen tissue samples and cultured cells……………………………………………….................48

2.1.17 Reagents for Global methylation assay using Imprint ® Methylated DNA quantification kit……………………………….....49

2.1.18 Reagents for Quantification of DNA methyltransferase activity using QPCR……...49

2.1.19 Human cancer cell line……………………………………..…………....**50**

2.1.20 EIAV, FIV and HIV vectors……………………….........................**50**

2.1.21 Compositions of buffers and solutions………………….............**50-51**

2.2 METHODS…………………………………………….52-77

2.2.1 Quantification of nucleic acids………………………………..……**52**

2.2.2 Agarose gel electrophoresis……………………………….........**52-55**

2.2.3 Preparation of LB agar plates and LB medium……………………....**53**

2.2.4 Transformation of PCR products…………………………………....**54**

2.2.5 Selection of transformed clones using 'blue/white screening'………..**55**

2.2.6 Growth of transformed plasmids in LB medium……………………....**55**

2.2.7 Storage and recovery of bacterial cells for sequencing analysis……...**55**

2.2.8 Plasmid DNA purification……………………………................**56**

2.2.9 Restriction digestion of purified plasmid DNA……………….....**56-57**

2.2.10 Colony PCR…………………………………………….............**57-58**

2.2.11 DNA sequencing………………………………………….................**58**

2.2.12 LAM PCR…………………………………………………………**59-65**

2.2.13 454 sequencing……………………………………….................**65**

2.2.14 RNA extraction, purification and cDNA synthesis for RT (reverse transcriptase) and QPCR (Real -Time Quantitative (Q) PCR) analysis……………………………………………………........**66-67**

2.2.15 CDNA synthesis for RT PCR…………………………………........**68**

2.2.16 RT PCR amplification reagents……………………………………**69**

2.2.17 CDNA synthesis for QPCR…………………………………………**69-70**

2.2.18 QPCR for gene expression analysis and VCN analysis.....................**70-72**

2.2.19 Microarray analysis for gene expression of murine cellular genes retrieved by LAMPCR and 454 sequencing and known E2F targets…………………………………………........**73**

2.2.20 Bioinformatics – STRING analysis.....73

2.2.21 Tissue culture.....73-74

2.2.22 Storage of cells derived from tissue culture.....74

2.2.23 DNA extraction from frozen tissue samples and cultured cells......74-75

2.2.24 Global methylation assay using Imprint ® Methylated DNA quantification kit.....75-76

2.2.25 Quantification of DNA methyltransferase activity using QPCR.....77

2.2.26 Bisulphite modification, methyl sensitive PCR and DNA sequencing.....77

2.2.27 Brief on n values and statistical methods used.....77

CHAPTER 3

3.1 RESULTS.....78-103

Investigation of vector associated geneotoxicity in mice following in utero delivery of EIAV, FIV and HIV vectors

3.1.1 Detection of viral vectors (EIAV, FIV and HIV) in infected mouse normal and tumour tissue.....78-79

3.1.1.1 Sequence comparison of vector sequences *in vivo*.....80

3.1.1.2 Comparison between VCN data obtained from Southern blotting and QPCR.....81-89

3.1.1.3 VCN measurements in EIAV infected mice.....90

3.1.1.4 VCN measurements in FIV infected mice.....91

3.2.2 Retrieval of proviral insertion sites and flanking genomic sequences by LAM PCR and BLAST analysis.....92-96

3.2.3 Analysis of the position and role of EIAV and FIV gene – gene ontology.....97-103

CHAPTER 4

4 INTRODUCTION……………………………..……..……104-108

4.1 Provirus integration can influence gene expression……………..…….**104-105**

4.1.1 Provirus and splicing mechanisms: the effect on gene expression.......**105**

4.1.2 Target site selection and gene expression patterns……....…........**106-107**

4.2 Proviral insertions and the construction of cancer gene pathways…….........**107**

4.2.1 Hepatocelllar Carcinoma (HCC)…………………………………...**108**

4.3 RESULTS………………………………………..109-134

Investigation of vector associated geneotoxicity in mice following in utero delivery of EIAV, FIV and HIV vectors

4.3.1 RT PCR analysis of tumour gene expression linking PLIONhAATGFP FIV derived vector insertions with alterations in gene expression…………………………………………........**109-113**

4.3.2 Real time PCR measurement of transcription levels of genes with FIV provirus insertions using QPCR analysis………..……..…..**114**

4.3.3 Q-PCR analysis in EIAV infected samples…………………...**115-120**

4.3.4 Analysis of FIV insertion gene pathways by 'String'……..….**121-123**

4.3.5 Analysis of EIAV insertion gene pathways using the STRING database and microarray analysis………………..….**124-131**

4.3.6 The effects of SMART2Z integration on normal splicing of the *Pah* gene……………………………………………........**132-134**

CHAPTER 5

5 INTRODUCTION……………………………………..134-147

5.1 The relationship between viral integration and E2F target genes……..........**134**

5.1.1 E2F and *Rb1* interaction is fundamental to E2F activity……...**135-136**

5.1.2 E2F and DNA damage……………………………………......**136-138**

5.2 Epigenetic modification of host genes following lentiviral vector delivery TSG methylation in HCC and non tumour liver……….....….138

5.2.1 Epigenetic regulation of transcription…………………………138-139

5.2.1.1 *Dnmt1*……………………………………………………….………140

5.2.1.2 *Dnmt3a* and *Dnmt3*b……………………………….……..140-142

5.2.1.3 Interaction between DNA methylation and chromatin……..142-143

5.2.2 Detection of DNA methylation changes………………………143-144

5.2.3 Methylation, mutagenesis and DNA damage…………………....144

5.2.4 The role of viral integration in methylation…………...………145-147

5.3 RESULTS………………………………………….……..148-169

Epigenetic modification and E2F regulation of host genes following lentiviral vector delivery

5.3.1 Microarray analysis of E2F targets *in vivo* to investigate the changes in E2F targets following lentivirus integration…………….…...148-149

5.3.1.1 E2F transcription factors and target genes are differentially expressed in EIAV infected tissues…………………...……150-155

5.3.2 Changes in DNA *methylase* expression in lentivirus infected tissues……………………………………….........156-161

5.3.3 Observations of global methylations associated with *in utero* application of lentivirus…………………………………………………162-163

5.3.4 Measurement of the effects of lentivirus infection on methylation *in vitro*……………………………………………………….....164

5.3.4.1 Global methylation levels: *In vitro* study…………………165-166

5.3.4.2 *Dnmt* methylation levels: *In vitro* study……………….....167-168

5.3.5 Analysis of VCN following the delivery of lentivirus vectors to liver cell *in vitro*……………………………………………….……….169

6 DISCUSSION………………………………..…….……...170-182

7 APPENDICES…………………………………………….......183-213

8 REFERENCES………………………………….……........214-259

9 BIBLIOGRAPHY AND ONLINE REFERENCES...……….260

LIST OF FIGURES

Figure 1. The genomic organisation of retroviruses……………………….**2**

Figure 2. A schematic overview of the retrovirus genome…………………**5**

Figure 3. A schematic overview of the retrovirus structure………….......**6**

Figure 4. The retroviral life cycle……………………………..………….**8**

Figure 5. Reverse transcription of the retroviral RNA genome…………….**9**

Figure 6. Integration of the retroviral DNA into the host cell genome.......**10**

Figure 7. Examples of lentivirus constructs used in gene transfer studies …………………………………………….......**16**

Figure 8. Production of retroviral vectors for gene therapy………………..**18**

Figure 9. Self inactivating (SIN) vectors…………………………….......**19**

Figure 10. Mechanisms of insertional mutagenesis associated with retroviral integration……………………….……………………**26**

Figure 11. Schematic representation of the pDRIVE vector used in transformation studies………………………………...…….…….**54**

Figure 12. A schematic overview of the steps required for LAM PCR…….**59**

Figure 13. Schematic representation of the vector constructs used for *in utero* studies……………………………...………………….…….**80**

Figure 14. Vector copy number versus age of death in EIAV infected samples…………………………………………….........**90**

Figure 15 Vector copy number versus age of death in FIV infected samples………………………………………………….**91**

Figure 16. Schematic image of FIV vector and oligo sequences used in LAM PCR analysis………………………………………………**94**

Figure 17. Proviral insertion sites retrieved from tumours by LAM PCR analysis………………………………………...……..**95**

Figure 18. Chromatagram files obtained from sequencing of LAM PCR products…………………………...…………..**96**

Figure 19. A typical agarose gel showing GAPDH expression in FIV treated mice………………………………………………........**109**

Figure 20. Relative gene expression levels of FIV insertions genes using RT PCR…………………………...**112-113**

Figure 21. Histograms used to represent gene expression data ($2^{-\Delta\Delta CT}$) obtained from FIV & EIAV treated mice using TaqMan analysis……………………………………………………..….**120**

Figure 22. STRING analysis used to predict protein-protein interactions between FIV insertion genes……………………………………**122-123**

Figure 23. STRING analysis used to predict protein-protein interactions between EIAV insertion genes……………………………...**125-129**

Figure 24. The relationship between *Uvrag* and *Park7* demonstrated using STRING analysis…………………………………………......**130**

Figure 25. IPA analysis: Interaction between *Park7* and *Uvrag*…….…….**131**

Figure 26. Representation of the *Pah* integration site and oligo sequences used to amplify the *Pah* gene…………………...……**132**

Figure 27. PCR analysis of splicing mechanisms of the *Pah* gene in EIAV treated mice………………………………………………....**133**

Figure 28. E2F mediated activation of cellular genes Dissociation of E2F: pRB complexes……………………………**136**

Figure 29. Schematic image representing the role of E2F in the DNA damage pathway…………………………………………….......**137**

Figure 30. Cytosine methylation……………………………………………...**139**

Figure 31. Cytosine methylation patterns in DNA mantainance and *de novo* methylation………………………………………….......**142**

Figure 32. Association between *Dnmt*, HDAC's, methyl binding proteins and methylated cytosines…………………………………….....**143**

Figure 33. Histogram representing *Dnmt* levels measured in FIV infected samples (normal liver………………………………....**157**

Figure 34. Histogram representing the relative gene expression values of *Dnmt's* in tumour samples infected with the FIV vector……………………………………………………………...**158**

Figure 35. *Dnmt* gene expression data obtained from EIAV infected mice (normal liver……………………………………….......**159**

Figure 36. *Dnmt* gene expression data obtained from EIAV infected mice (tumours)……………………………………………….**160**

Figure 37. *Dnmt* gene expression data obtained from HIV infected normal liver samples……………………………………………….**160**

Figure 38. DNA methylation levels in normal tissues associated with lentivirus transduction……………………………………....**162**

Figure 39. DNA methylation changes in hepatocellular carcinoma tissues transduced with lentivirus vectors……………………………….**163**

Figure 40. Global methylation levels in uninfected, lentivirus treated and retrovirus treated HepG2 cells at 6hrs, 24hrs and 30 hrs post infection……………………………………………………….**163**

Figure 41. *Dnmt3a* levels seen in negative (uninfected) and vector treated (MLV, FIV, HIV and EIAV) HepG2 cells……………..**167**

LIST OF TABLES

Table 1. The retrovirus family members: Complexity and oncogenic potential……3

Table 2. Basic components of the retrovirus genome……5

Table 3. Components of the lentivirus genome……12-13

Table 4. General chemicals and reagents……41

Table 5. Summary of reagents required for transformation protocol……41

Table 6. Summary of reagents used in LAM PCR analysis……42

Table 7. List of primers used for LAM PCR analysis……43

Table 8. Reagents used for isolation of total RNA from tissue samples……44

Table 9. General reagents used for *DNase I* treatment……44

Table 10. General reagents used for CDNA synthesis of samples used for RT PCR……44

Table 11. General reagents used for RT PCR……44

Table 12. Oligonucleotides used RT PCR amplifications……45

Table 13. General reagents/chemicals used for QPCR reactions……45

Table 14. General reagents/chemicals used for CDNA synthesis of total RNA for QPCR reactions……45

Table 15. General reagents/chemicals used for TaqMan PCR reactions……46

Table 16. TaqMan primer probe sets used for VCN studies……46

Table 17. TaqMan probes and unlabelled oligonucleotide sequences used for TaqMan analysis of lentivirus insertions genes……47

Table 18. General reagents used for tissue culture……48

Table 19. Reagents used for DNA extraction of frozen tissue and cultured HepG2 cells lines……48

Table 20. General reagents used for global methylations assays……49

Table 21. TaqMan gene expression assays used to quantify DNA methyltransferase activity……49

Table 22. Reagents /chemicals required for preparation of LB agar plates and LB medium……53

Table 23. Reagents used for restriction digestion of plasmid DNA

using *EcoRI*……………………………………………………….57

Table 24. Reagents and volumes required for colony PCR…………………57

Table 25. PCR parameters used for colony PCR……………………..…..58

Table 26. LAM PCR: Reagents and volumes required for linear PCR…....60

Table 27. PCR parameters required for linear PCR reactions………………60

Table 28. LAM PCR: Reagents and volumes required for second strand synthesis……………………………………………………...61

Table 29. LAM PCR: Reagents and volumes required for restriction digestion using *Tsp509I*…………………………………………….62

Table 30. LAM PCR: Reagents and volumes for preparation of the linker cassette using the ligase enzyme…………………………..63

Table 31. LAM PCR: Reagents and volumes required for first round PCR……………………………………………………....64

Table 32. PCR parameters required for first round PCR……………………64

Table 33. LAM PCR: Reagents and volumes required to make a second round PCR master mix ………………………………………......65

Table 34. Preparation of a reaction mixture for *DNase I* treatment………..66

Table 35. Preparation of a RNA primer master mix for RT PCR……………68

Table 36. Preparation of a cDNA synthesis master mix for RT PCR……….68

Table 37. Preparation of RT PCR master mix………………………………..69

Table 38. General reagents/chemicals used for CDNA synthesis of total RNA for QPCR reactions……………………………………70

Table 39. PCR parameters used for cDNA synthesis using the MultiScribe Reaction kit……………………………………………..……70

Table 40. Reagents/chemicals used for preparation of TaqMan PCR mastermix for amplifications……………………………………70

Table 41. PCR parameters used for TaqMan PCR parameters used for TaqMan amplification reactions…………………………………...71

Table 42. Animal tissues subjected to QPCR to determine VCN using primer/probes…………………………………………..82-83

Table 43. Comparison between VCN data obtained from Southern blotting and QPCR…………………………………………………….88-89

Table 44. Vector integration sites in FIV and EIAV infected Tumours……………………………………………………..…**99-103**

Table 45. Relative gene expression data obtained for FIV insertion genes using TaqMan analysis……………………………………..…...**114**

Table 46. Relative gene expression data obtained for EIAV insertion genes using TaqMan analysis…………………………………….**117-119**

Table 47. E2F family members…………………………………………....**135**

Table 48. Changes in the expression of E2F target genes in EIAV infected liver and tumour samples………………………………..**150-155**

Table 49. *Dnmt* expression in mouse tissue samples after FIV infection……………………………………………………..**157**

Table 50. *Dnmt* gene expression data obtained from EIAV infected mice……………………………………………...**159**

Table 51. Expression of *Dnmt* genes in HIV infected samples………....**161**

Table 52. Global methylation levels *in vitro*………………………………**166**

Table 53. *In vitro* study: *Dnmt* gene expression represented as relative methylation levels ($2^{-\Delta\Delta CT}$)…………………………....**168**

Table 54. VCN data obtained from lentivirus treated mice using QPCR analysis……………………………………………………**169**

ABBREVIATIONS

Abs	Absorbance
ADA	adenosine deaminase
AIDS	acquired immune deficiency syndrome
ALV	Avian leukosis virus
AML	acute myeloid leukaemia
A.N	Adame normal liver
ASLV	Avian sarcoma leukosis virus
A.T	Adame tumour liver
ATP	adenosine triphosphate
βME	beta-mercaptoethanol
BLAST	basic local alignment search tool
bp	base pair
BSA	bovine serum albumin
CA	capsid protein
CAEV	Caprine arthritis-encephalomyelitis
CD	cluster of differentiation
CD34+	haematopoietic progenitor cells
cDNA	complementary deoxyribonucleic acid
CGD	chronic granulomatous disease
CH3	methyl group
Chr	chromosome identity
CHS4	chicken β globulin 5'DNase I hypersensitive site-4
CIS	common insertion site
CMV	Cytomegalovirus
CpG	cytosine followed by guanine
cPPT	central polypurine tract
CT	cycle threshold
ΔLNGFR	truncated version of p75 neurotrophin receptor
dAdo	2'-deoxyadenosine
dH_2O	distilled water
dLNGFR	truncated nerve growth factor receptor
DMEM	Dulbecco's modified eagle's medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNMT	DNA methyl transferase
dNTPs	deoxynucleotide triphospates
dsDNA	double stranded DNA
DTT	Dithiothreitol
DU/dUTPase	deoxyuridine triphosphatase
EB	elution buffer
EDTA	ethylene diamine tetraacetic acid
EIA	equine infectious anaemia
EIAV	equine infectious anaemia virus
EMA	European medicines agency
Env	viral envelope
EVI1	ecotropic viral integration site
FDA	US food and drug administration
FIV	feline immunodeficiency virus

FIX	clotting factor IX
FOC	Flora ovarian cyst
GAG	group specific antigen
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
γc	gamma chain
GFP/eGFP	green fluorescent protein
G.N	Gerry normal liver
GO	gene ontology
GP	Glycoprotein
G.T	Gerry tumour liver
GVHD	graft versus host disease
HDAC	histone deacetylases
hrs	Hours
HBV	hepatitis B virus
HBx	hepatitis B virus X protein
HCV	hepatitis C virus
HCC	hepatocellular carcinoma
HepG2	hepatocellular carcinoma cell line
hFIX	human clotting factor IX
HIV	human immunodeficiency virus
HPRT	hypoxanthine-guanine phosphoribosyltransferase
HR'SIN-CPPT-S-FIX-W	HIV vector construct containing SFFV promoter
HSC	haematopoietic stem cell
HSV	herpes simplex virus
HTLV	human T-cell leukaemia virus
ICF	immunodeficiency centromeric instability and facial anomalies
IL2	interleukin 2
IL2RG	interleukin 2 receptor, gamma
IM	insertional mutagenesis
IMS	industrial methylated spirit
IN	Integrase
IPA	ingenuity pathway analysis
IPTG	isopropyl β-D-1-thiogalactopyranoside
IU	infectious unit
J.N	Jonas normal liver
J.T	Jonas tumour liver
kb	kilobase pairs
LacZ	β-galactosidase gene
LAM-PCR	linear amplification mediated polymerase chain reaction
LB	Luria-Bertani
LCMV	lymphocytic choriomengitis virus
LEDGF	lens epithelium-derived growth factor
LTR	long terminal repeat
LV	Lentivirus
MA	matrix protein
MDR1	multidrug resistance protein 1
$MgCl_2$	magnesium chloride
MGI	mouse genome informatics
MHC	major histocompatibility complex

MLV	murine leukaemia virus
MOI	multiplicity of infection
Mo-MLV	Moloney murine leukaemia virus
mRNA	messanger ribonucleic acid
MVV	Maedi-visna virus
NC	Nucleocapsid
NEF	negative regulatory factor
NC	Nucleoplasmid
NK	natural killer cell
OC	ovarian cyst
ORI	orientation of vector into RefSeq gene
OTC	ornithine transcarbamylase deficiency
P1	re-suspenion buffer
P2	lysis buffer
P3	neutralising buffer
PB	binding buffer
PBL	peripheral blood lymphocytes
PBS	primer binding site
PBS^	phosphate buffer soloution
PCR	polymerase chain reaction
PE	wash buffer
PEG	polyethylene glycol-conjugated bovine
PIC	preintegration complex
pLIONhAATGFP	FIV construct containg eGFP reporter gene
POL	Polymerase
PPT	polypurine tract
PR	Protease
PSI/Ψ	packaging signal
QPCR	quantitative real time polymerase chain reaction
R	repeat region
RB	Retinoblastoma
RCL	replication competent lentivirus
RCR	replication competent retrovirus
RCV	replication competent virus
RDV	replication defective retrovirus
RefSeq	reference sequence
REV	Regulatory of virion
RIS	retroviral insertion site
RLGS-M	restriction landmark genomic scanning for methylation
RNA	ribonucleic acid
RPE	retinal pigment epithelial cells
rpm	rotations per minute
RQ	relative quantification level
RRE	rev response element
rRNA	ribosomal ribonucleic acid
RT	reverse transcriptase
RTCGD	retroviral tagged cancer gene database
RT-PCR	reverse transcription polymerase chain reaction
RV	Retrovirus
SA	splice acceptor

SCID	severe combined immunodeficiency
SD	splice donor
SEM	standard error of the mean
SFFV	spleen focus forming virus
SIN	self inactivating
SIV	simian immunodeficiency virus
SMART2hFIX	EIAV vector construct containing factor IX transgene
SMART2WΔ	EIAV vector construct containing mutated WPRE sequence
SMART2Z	EIAV vector construct containing lacZ transgene
SpT	spontaneous tumour liver
STRING	search tool for the retrieval of interacting genes/proteins
SU	Surface subunit
TAR	transactivator responsive element
TAT	transcriptional activator
TBE	tris borate acid
TE	tris-EDTA
TM	transmembrane subunit
tRNA	transfer ribonucleic acid
U3	unique 3'
U5	unique 5'
UGT1A1	UDP-glucuronyltranferase 1A1
VCN	vector copy number
VPR	viral protein R
VPU	viral protein U
VPX	viral protein X
VSV-G	vesicular stomatitis virus G glycoprotein
WHO	world health organization
WPRE	woodchuck hepatitis virus posttranscriptional regulatory element
X-Gal	5-bromo-4-chloro-3-indolyl β-D-galactoside
X-SCID	x-linked severe combined immunodeficiency
18SRNA	18S ribosomal RNA

ACKNOWLEDGEMENTS

Firstly I would like to thank my supervisor Dr. Michael Themis for contributing to my personal and academic growth over the past four years. Under your supervision I have learnt a lot about what it means to be a scientist and the value of good research. I would also like to thank Dr. Rhona Anderson, my second supervisor for overseeing my studies.

I would like to acknowledge the generosity of several collaborators; in particular Dr. Manfred Smidth and Dr. Robert Andrews for their contribution to 454 sequencing and Microarray analysis described in Chapter 5.

I would like to thank Brunel's Biosciences department for making it a great place to work in but in particularly several members who gave me practical advice, training and reagents when needed including Dr. Mark Pooks group, Mr. Derek Fisher, Ms. Denise May, Ms. Helen Cox, and Ms.Christine Newton. I would particularly like to acknowledge Ms. Safia Reja for her help with many things including the maintenance of the HepG2 cells (Kindly provided by Dr. Amanda Harvey).

I am grateful to Drs. Laurence Bugeon, Terry Roberts and Helen Foster for their technical advice on QPCR analysis.

I am very thankful for the contribution of Dr. Evgeny Makarov for his advice and training on all things splicing and Ms Romana Linkova for her contribution to IPA analysis.

I am especially grateful to Dr. Matthew Themis for training me on several techniques including LAM PCR and his advice on all things that fall under the umbrella of life.

I have met some great people who have shared this era of my life and kept me grounded. Thank you especially Ms Elena Zachariades and Ms. Safia Reja for your friendship and support.

Thank you to my friends outside of Brunel who despite my absence still encouraged me to stay focused.

Last but not least I would like to acknowledge my wonderful family – Dad, Mum, Jessica, Benedicta, Edwin and Emmanuel. You have been my constant support system throughout my studies. Without you all I would have lost my mind. Thank you for believing in me, helping me to laugh when things got tough and picking me up from various locations when my car got tired of doing its job.

DECLARATION

I, Emma Awele Osejindu, hereby declare that the work presented in this thesis was carried out by myself unless otherwise stated.

1.1 Gene therapy

Current statistics indicate that 13,000 births a year in the UK are affected by congenital or genetically determined abnormalities (http://www.geneticalliance.org.uk/education3.htm).

Gene therapy has emerged as a promising way to treat genetic diseases by introducing a ‘normal’ version of a gene into patient cells to modify host gene expression and reach a therapeutic effect. For some diseases stem cell therapy is possible, however, the uncertainty and risk of complications associated with HLA-mismatched bone marrow transplantation means treatment using stem cells in a autologous manner is important. Hence, gene therapy may be used either on patient stem cells or somatically.

Clinical studies conducted during the 1960’s demonstrated the ability of foreign DNA to penetrate mammalian cells and become expressed (Borenfreund and Bendich, 1961; Friedman, 1992). In these early studies, Borenfreund and Bendich (1961; Bendich, 1961; Rieke, 1962; Chorąży *et al*, 1963; 1964) isolated and labeled DNA from pneumococci and human leucocytes with tritium before adding the DNA to growing cultures of HeLa cells. Using autoradiography to measure radioactivity in the treated cells it was demonstrated that gene transfer of exogenous DNA was possible and that incorporation into the nucleus of mammalian cells occurred within 6-24 hours post treatment. Once the genetic code was the deciphered between 1961 and 1966 (Nirenberg, 2004; Nienhuis, 2008) interest in genetic engineering and gene transfer methods has mounted and made possible the characterisation and delivery of therapeutic genes. More recently, gene delivery has been improved by using viruses as vectors. One such virus is the RV and understanding retroviral biology has facilitated and promoted the use of these viruses to carry out gene transfer in a more efficient way. In order to discuss the use of these viruses in gene therapy an overview of retroviral biology is firstly presented.

1.1.1 Retroviruses an overview

The term retroviruses (RV) are used to describe a large family of RNA viruses with similar structures (Figure 1).

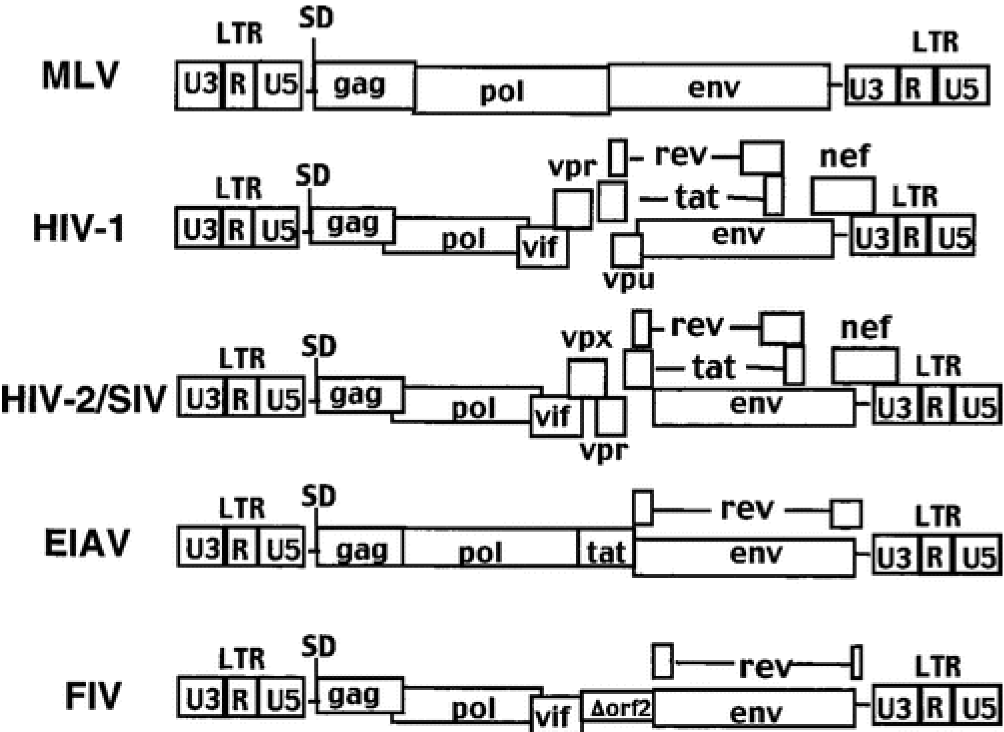


Figure 1. The genomic organisation of murine leukaemia virus (MLV), Human immunodeficiency virus type 1 (HIV-1), Human immunodeficiency virus type 2 (HIV-2), Simian immunodeficiency virus (SIV), Equine infectious anaemia (EIAV) and Feline immunodeficiency virus (FIV) are shown above.
Several similarities between the RV's have been found. These include the long terminal repeat (LTR), the group specific antigen (gag) gene, the polymerase gene (pol) and the gene coding for the viral envelope (env). The *LTR*, *gag*, *pol* and *env* genes play important roles in generating viral particles by producing an RNA genome that drives gene expression. The RT enzymes are used to convert the genome into DNA and integrate it into the host genome. *Env* genes produce virus glycoproteins which are needed to coat the virus with glycoprotein receptors to allow infection of new host cells. The main difference between RV's is the presence of the accessory gene Regulatory of virion *(*rev) in HIV, SIV, EIAV and FIV and not MLV genomes. Additional accessory genes: Transcriptional activator (tat), Negative regulatory factor (nef) and Viral protein R (vpr) exist in several RV's but not the MLV genome. The HIV-1 and HIV-2/SIV are the only RV genomes that contain a Viral protein U (vpu) gene. In addition, Viral protein X (vpx) is found in SIV and HIV-2 but not in HIV-1. Image taken from Tang, Kuhen and Wong-Staal, (1999).

RV particles are typically 80-100nm in size and consist of RNA encapsulated by a lipid envelope and viral glycoproteins (Coffin, Huges and Varmus, 1997). The RV genome contains two single, linear strands of RNA which possess positive polarity (Coffin, 1996). Positive stranded viral RNA has the same sequence as mRNA. Hence these viruses can use their RNA genome directly as mRNA that can be translated by the host cell (Levinson, 2006). The defining feature of RV's is its replicative strategy which does not result in death of the host cell (Coffin, 1979). RV's transcribe their RNA into a complementary DNA (cDNA) that can be integrated into the host genome following infection via a viral encoded protein called the integrase. This ability to alternate genetic material between RNA in the virion and DNA in infected cells is a crucial feature of the viral life cycle (Temin *et al*, 1993). The RV family can be subdivided into seven groups (Table 1) according to their nucleotide sequence composition and distinctive features which can be visualised by transmission electron microscopy (Coffin, Huges and Varmus, 1997; Brenner and Malech, 2003).

Number	Genus	Example	Genome	Oncogenic potential
1	Avian - leukosis sarcoma viral group	Rous sarcoma virus	S	O
2	Mammalian – B type viral group	Mouse mammary tumour virus	S	O
3	Murine leukaemia related viral group	Moloney murine leukaemia virus	S	O
4	D-type viral group	Mason – Pfizer monkey virus	S	O
5	Human T cell leukaemia-bovine leukaemia viral group	Human T ell leukaemia Virus	C	O
6	Spumaviruses	Foamy virus	C	
7	Lentiviruses (LV)	Feline Immunodeficiency virus	C	

Table 1. RV family. Simple RV's and complex RV's are represented as orange circles and blue circles, respectively. Note RV groups can display oncogenic potential and these are marked by red stars and known as oncoRVes.

Studies conducted using murine leukaemia virus (MLV) during the 1970s contributed greatly to our current knowledge of RV infection and integration mechanisms (Baltimore, 1970; Temin, 1980; Miller *et al*, 1990; Miller, 1997; Daly and Chernajovsky, 2000; Kay *et al*, 2001; Anderson and Hope, 2005). Many RV's contain oncogenes (i.e. *myb, v-myc and ras*) that were believed sequestered over several rounds of their life cycle involving integration and then generation of new particles that became responsible for the initiation and maintenance of neoplastic transformation of target cells (Spector, Varmus and Bishop, 1978; Sheiness and Bishop, 1979; Gonda, Sheiness and Bishop 1982; Bishop, 1983; Varmus, 1984; Vanin *et al*, 1994). RV's which possess oncogenic potential are collectively known as oncoRVes (Table 1). RV's are further classified into two groups called complex and simple viruses according to their genomic structures (Table 1). The RV genome common to all RV's consists of three coding genes called *gag, pol* and *env* required for viral replication of simple and complex RVes (Figures 1 and 2). The *gag* gene encodes the structural proteins that form the structural matrix, capsid and nucleoproteins found in the viral core. The *pol* gene yields the viral enzymes reverse transcriptase, RNase H and integrase responsible for cDNA production and integration, respectively. The *env* gene encodes the glycoproteins that protrude from the lipid bilayer surrounding the viral particle (Figure 3). These proteins are used to attach virus particles to receptors on the surface of host cells and determine the viral tropism (Brenner and Malech, 2003; Peifer and Verma, 2001). The RV genome is flanked by duplicated regions in the 5' and 3' ends termed Long Terminal Repeats (LTR) (Temin, 1993). The LTR is composed of U3 (unique 3), R (repeat) and U5 (unique 5) sequences required for viral gene expression and production of new genomes (Figure 1). Table 2 provides details of the genes carried by RV's.

Table 2. Basic components of the RV genome

Gene	Function
Gag	Production of Gag polyproteins required for the construction of RV particles. Production of viral matrix, capsid and nucleocapsid proteins.
Pol	Encodes the enzymes *protease, reverse transcriptase* (*rt*) and *integrase* enzymes which are required for cleavage of viral polyproteins into mature proteins; replication of viral RNA into DNA and integration of viral DNA into the host genome respectively.
Env	Encodes the viral glycoproteins present on the surface of the virus particle. The envelope precursor molecule breaks down into the surface glycoprotein and transmembrane structures.
LTR	Drives viral gene expression required for replication, reverse transcription and production of new viral genomes. Composed of 3 essential sequences: U3, R and U5.
Repeat - R	Essential for reverse transcription and replication. Also provides polyadenylation signal in MLV and interacts with a transaactivation response region in HIV-1.
Unique 3 - U3	Contains essential promoter and transcriptional enhancer sequences.
Unique 5 - U5	Consists of sequences required for initiation of reverse transcription.

Three major coding domains present in the wild type configuration of all RV vectors are shown. These genes are required for successful gene transfer and completion of the viral life cycle.

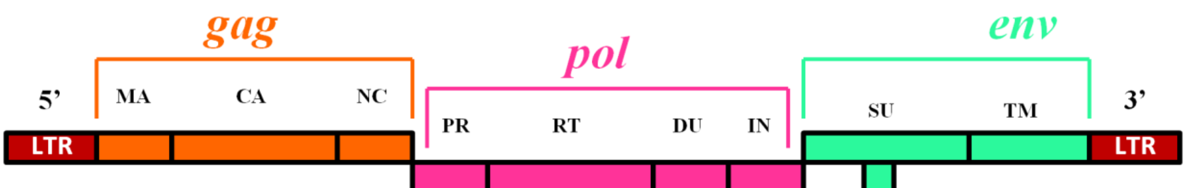


Figure 2. A schematic overview of the RV genome. The viron RNA is typically 7-12Kb in size. The genomic structure of RNA consists of the *gag* gene located in the 5' end of the viral genome followed by the *pol* and *env* genes. The RNA genome is flanked by the LTR regions. The *gag* gene encodes the matrix protein (MA), capsid protein (CA) and nucleoplasmid (NC). The *pol* gene encodes reverse transcriptase (RT), protease (PR), integrase (IN) and deoxyuridine triphosphatases
(dUTPases). The *env* gene encodes the surface subunit (SU) and transmembrane subunit (TM).

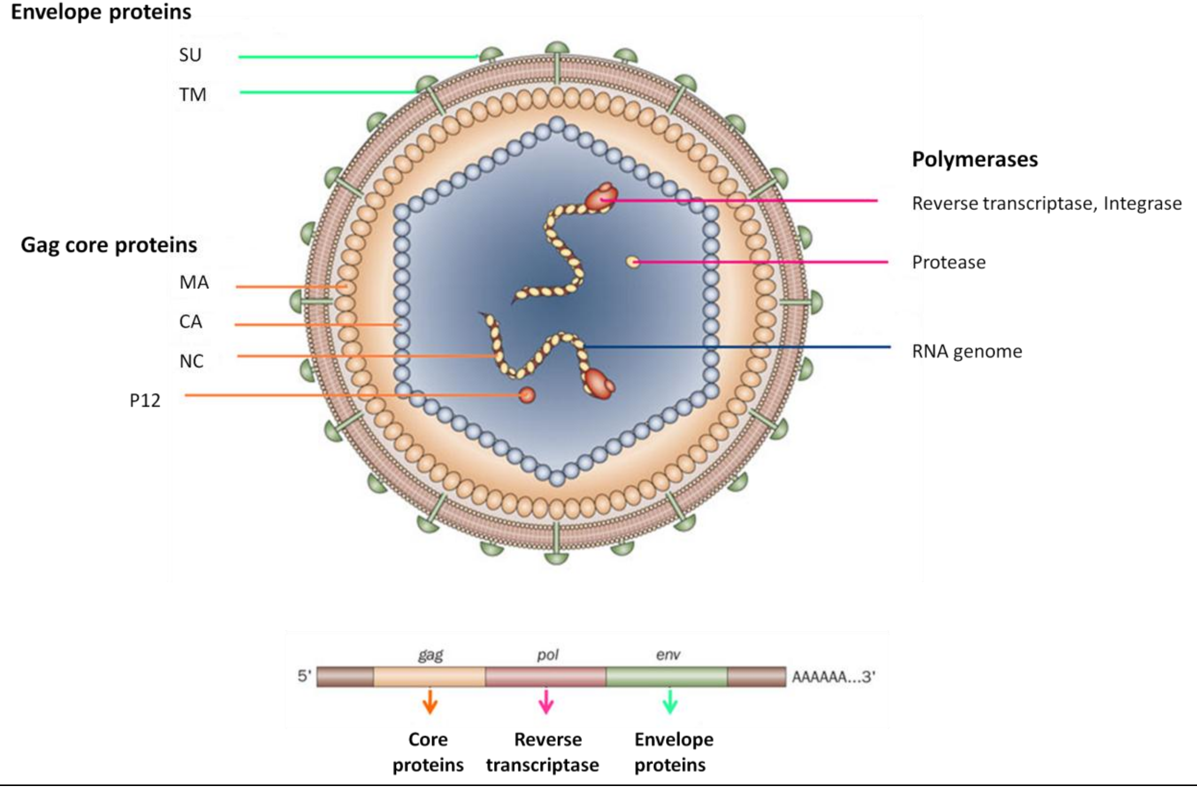


Figure 3. A schematic overview of the RV's structure. The RV particle is composed of an RNA genome packaged in capsid proteins and surrounded by a lipid envelope. The RNA genome consists of two single stranded RNA molecules. The *gag* gene encodes several *gag* core proteins. The *pol* gene codes for the production of enzymes reverse transcriptase, protease and integrase. The *env* gene provides the genetic code for envelope glycoprotein structures. These glycoprotein structures are deposited in the lipid bilayer of the cell and surround the virus particle after budding of the virus from the host cell membrane. The viral particle is typically 100nm in size. A modification of an image created by Silverman *et al*, (2010).

1.1.2 The RV life cycle

Cell binding and entry steps

The life cycle of a RV is shown in Figure 4 and begins with attachment of the virus to the host cell membrane via specific virus ligand/ host receptor interaction.
The virus envelope protein recognises specific host cell receptors (for example CD4, CD9 and Rec1) and uses cell surface proteins/molecules to facilitate entry into target cells (Miller, 1996). The envelope protein is encoded as a polyprotein precursor which is cleaved into a surface unit (SU) and a transmembrane (TM) domain (Overbaugh *et al,* 2001). The SU protein is located on the outside of the virus particle and the TM protein anchors the SU to the viral membrane. The SU protein initiates entry of the RV particle by binding to specific cell surface proteins (Overbaugh *et al,* 2001). The SU molecule then induces a conformational change that exposes fusion proteins present in the TM (Cosset *et al*, 1995; Damico and Bates, 2000). Exposure of the viral core facilitates interaction between viral and specific host cellular proteins via direct fusion or endocytosis. This process depends on the virus envelope ligand and cell receptor used for entry. Binding of the HIV gp120 envelope glycoprotein to the CD4 structures in host cells induces conformational changes in the HIV gp120 glycoprotein. These changes results in the exposure of specific chemokine receptors including CCR5 and CXCR4 that allows virus entry (Kwong *et al*, 1999). In contrast Tva serves as the cellular receptor for the RV subgroup A avian sarcoma and leukosis virus (ASLV-A). The interaction between Tva and the ASLV-A specific protein EnvA has been well characterised by several studies (Bates *et al*, 1993; Mothes *et al*, 2000; Wang *et al*, 2001). Several groups have also reported the importance of pH conditions in RV cell binding. RV's such as HIV-1, and MLV fuse with the plasma membrane of host cells in a pH independent manner (Wang *et al*, 2001). In contrast RV's such ASLV-A are described as pH dependent viruses that require low pH conditions to proceed with receptor mediated endocytosis (Kizhatil and Albritton, 1997; Hernandez *et al*, 1997; Damic and Bates, 2000; Katen *et al*, 2001; Wang *et al*, 2001). After entering the cell membrane the retroviral nucleoprotein core particle is delivered into the cytoplasm of the cell.

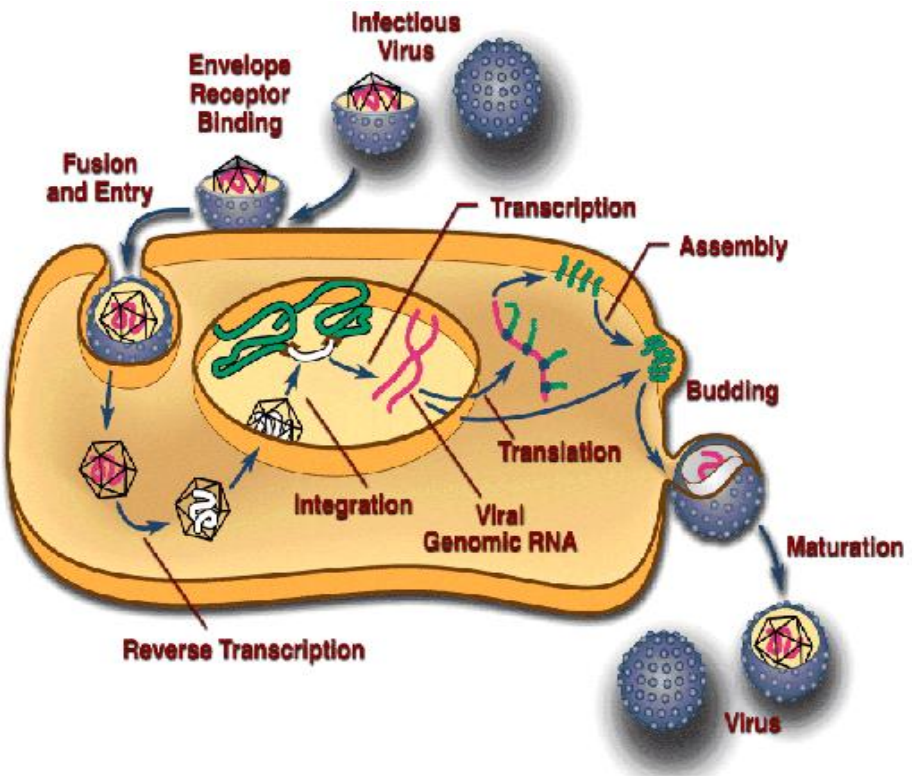


Figure 4. The viral life cycle begins with the infection of a host cell. Cell entry is facilitated by fusion of the virus with the host cell membrane. The virus particle uncoats and deposits its proteins into the cytoplasm of the target cell. CDNA synthesis (reverse transcription) and integration of viral cDNA are catalysed by viral polymerases including reverse transcriptase. After integration virus proteins are assembled and viral particles are released from the host cell via budding pathways. Image taken from ncbi website: http://www.ncbi.nlm.nih.gov/RVes/

Reverse transcription

The production of linear double stranded DNA from RNA molecules by reverse transcription is a pre-requisite for the integration of the double stranded DNA (provirus) into the host cell nuclear DNA. The retroviral nucleoprotein containing two copies of (+) viral RNA is uncoated in order to allow reverse transcription to occur (Figure 5 and 6).

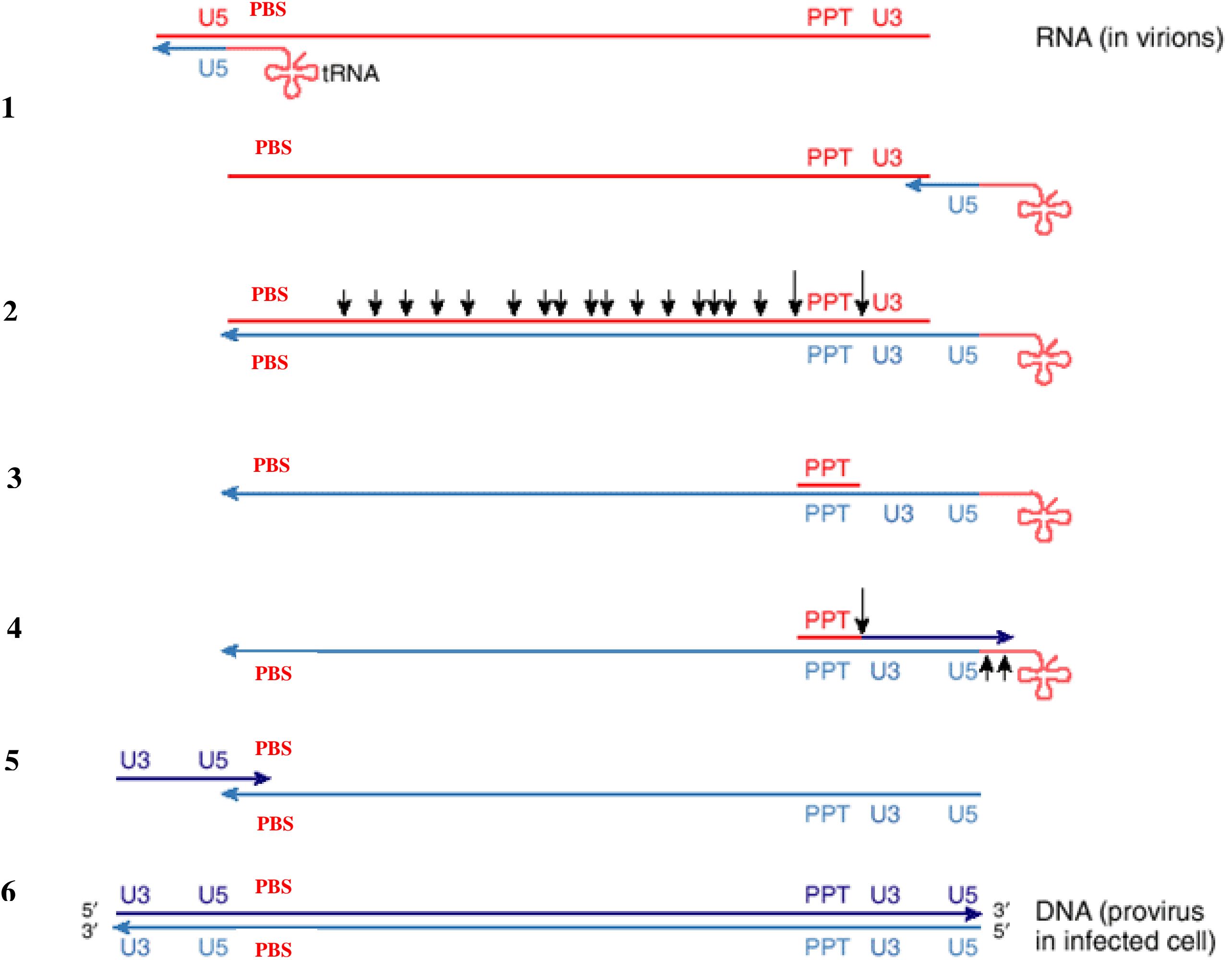


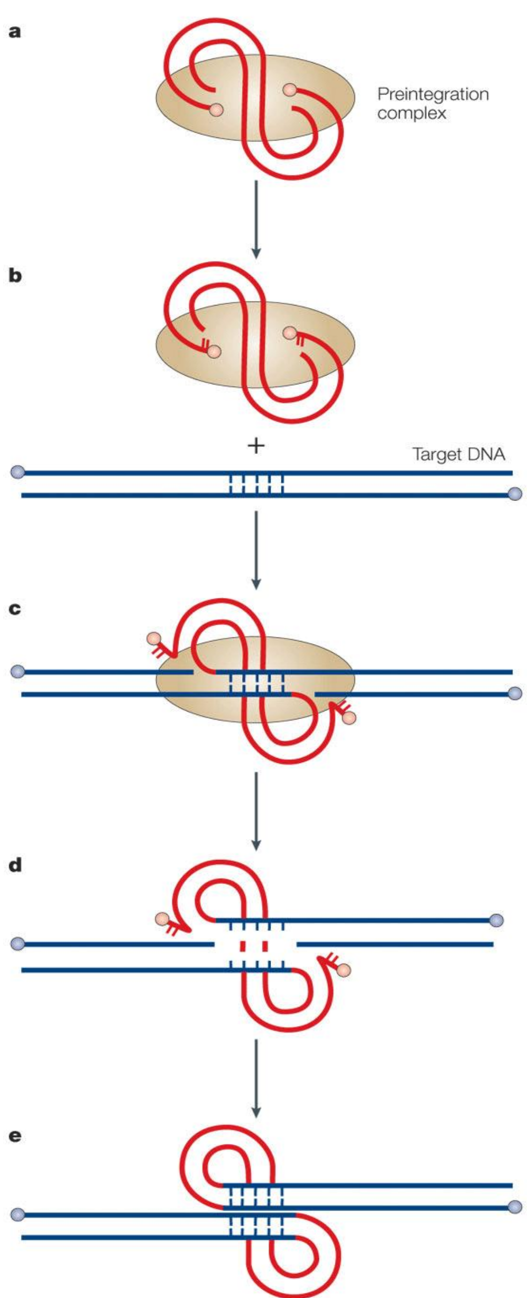
Figure 5. The overall mechanisms involved in reverse transcription of the viral RNA genome to produce a linear DNA duplex in RVes. The polypurine tract (PPT) and the primer-binding site (PBS) are shown above. 1. Priming by cellular tRNA is followed by initiation of minus-strand DNA synthesis (light blue). 2. RNA template (excluding the PPT region) is cleaved by the viral enzyme RNase H. Cleavage may be non specific (short arrows) or highly specific (long arrows) 3. Plus strand synthesis (dark blue). 4. RNase H cleavage of RNA template and removal. 5. The plus strand DNA (U3 –U5) is used as a template for the completion of plus strand synthesis. 6. DNA duplex (provirus) used for integration into the host genome of the target cell. The double stranded DNA contains duplications of the U3 and U5 regions at both ends. Image taken from Johnson, (2007).

Nuclear entry

Proviral DNA is assembled with other viral and cellular proteins in a pre-integration complex (PIC). It is thought that PIC interacts with the microtubule networks to allow transportation towards the nucleus. The PIC's of lentiviral complexes can be

transported across the nuclear pore without the breakdown of the nuclear envelope (Miller, Adam and Miller, 1990; Nisole and Saib, 2005). The *cis*-acting element cPPT plays a key role in assisting nuclear import of lentiviral PIC's in non-dividing cells (Skasko and Kim, 2008). In comparison, simple retroviral complexes such as MLV are dependent on the breakdown of the nuclear envelope during mitosis before transportation of PIC's to the cytoplasm can occur.

Integration



Integration of reverse transcribed retroviral DNA into the host cell genome is classified into 5 key steps Figure 6 a-e. a) The PIC composed of a double cDNA template (red lines). b) The integrase enzyme removes two nucleotides from the 3' end of viral DNA. This exposes recessed hydroxyl (OH) groups at the 3' end. c) Recessed OH groups are attached to host cell DNA via transestrification reactions. d) Base pairing between gaps are caused by unpairing between host and viral DNA junctions. e) Gaps are repaired resulting in integrated proviral DNA. Image taken from Bushman *et al*, 2008.

Viral gene expression, particle assembly and budding

Transcription of the provirus generates full length and spliced products. Full length viral transcripts may be packaged into new viral particles using a RNA sequence known as the packaging signal (Ψ). Alternatively full length viral transcripts also serve as templates for translation of viral proteins. Viral transcripts are then assembled at the cell periphery and released by budding off the plasma membrane.

1.1.3 Lentiviruses

Lentiviruses are a family of RVes described as agents of slow disease syndromes due to their long latency before clinical signs of chronic disease manifestation (Narayan and Clements, 1989). Lentiviral infection can cause disease in sheep, goats, cats, cattle, horses, monkeys and human species due to viral infection with Maedi-Visna virus, (MVV) caprine arthritis-encephalomyelitis virus (CAEV), feline immunodeficiency virus, (FIV) bovine immunodeficiency virus, equine infectious anemia virus, (EIAV) simian immunodeficiency virus, (SIV) and human immunodeficiency virus (HIV), respectively (Narayan and Clements, 1989; Clements and Zink, 1996). Lentiviral infection is associated with various pathogenic diseases including degeneration of haematological systems (EIAV), neurological systems (HIV and FIV), multi-organ failure and cancer (HTLV) which can ultimately result in death (Tang *et al*, 1999). The lentivirus genome is complex due to the presence of additional genes that contribute to efficient virus replication and infection (Pfeifer and Verma. 2001; Brenner and Malech, 2003). Regulatory genes (*tat* and *rev*) and auxiliary ('accessory') genes (*vif, vpr, vpu* and *nef*) are absent in simple RV such as MLV (Table 3; Pfeifer and Verma, 2001). A representation of the lentivirus genome is presented in Figure 7 and differences and similarities between the genes carried by these viruses are shown in Table 3.

Table 3. Components of the lentivirus genome

Lentiviral gene	Function	Simple RV	Lentivirus
Structural genes			
Gag	See Table 2	√	EIAV FIV HIV
Pol	See Table 2	√	EIAV FIV HIV
Env	See Table 2	√	EIAV FIV HIV
LTR (R,U3,U5)	See Table 2	√	EIAV FIV HIV
Regulatory genes			
Tat	Involved in activation of LTR during viral replication stages. Increases initiation and elongation of viral RNA transcription. Binds to the TAR region of viral RNA and to host proteins.		EIAV HIV
Rev	Promotes nuclear transport and translation of unspliced and partially spliced viral transcripts across nuclear membrane. Interacts with transcripts containing rev responsive element (RRE).		EIAV FIV HIV
Accessory genes and elements			
Vif	*Vif* increases infectivity and pathogenicity of HIV-1 and FIV. Increases infectivity also stabilises the pre-integration provirus. *Vif* also targets the APOBEC3 family of proteins which act as inhibitors of retroviral replication.		FIV HIV
Vpr	Aids infectivity and pathogenesis of HIV-1. Binds to nuclear pore complex and assists infection in non-dividing cells.		HIV
Vpu	Functions in infectivity and virulence of HIV-1. Enhances the assembly and release of virions from human cells following infection. Triggers CD4 degradation.		HIV
Nef	Involved in viral replication where it facilitates the ability of virus to evade host defense and reverse transcription. Down regulates the CD4 receptor.		HIV
Packaging signal Ψ	Sequence positioned between the ATG codon within *gag* and the major splice donor site. Essential for transgene expression and encapsidation mechanisms.	√	EIAV FIV HIV

The lentiviral superfamily are characterised by their complex genomes. Several genes are contained in the wild type configuration of all lentivirus vectors. Accessory genes facilitate the steps required for successful gene transfer and completion of the viral life cycle. Several auxiliary genes are species specific. HIV2 and EIAV differ from other lentiviral counterparts as they lack the *vpu* gene. In addition FIV lacks the *Tat* gene although it contains a putative *Tat* gene called *Orf2*. Regulator and accessory proteins are often modified or eliminated from lentiviral constructs used for gene transfer.
The presence of genes in the genome of simple and complex lentiviruses are represented by black ticks. The presence of genes in EIAV, FIV and HIV are represented as brown, pink and blue circles, respectivley.

1.1.3.1 Human immunodeficiency virus (HIV)

HIV has been extensively studied since the occurrence of the acquired immune deficiency syndrome AIDS epidemic in the 1980's. The HIV genome is approximately 10kb in size and exists in two forms known as HIV-1 and HIV-2. HIV-1 primarily infects CD4 T cells and macrophages (Delassus *et al*, 1991). HIV infection is a major cause of morbidity and mortality in many parts of the world. According to global reports produced by the world health organisation (WHO) in 2009, 33.3 million people were living with HIV. Out of this population 15.9 million were women http://www.who.int/hiv/data/2009_global_summary.png. A substantial proportion of those infected with HIV develops acquired immune deficiency syndrome (AIDS) and ultimately die. 1.8 million people died as a result of HIV infection in 2009 http://www.who.int/hiv/data/2009_global_summary.png. Research into HIV has allowed a detailed understanding of the HIV life cycle and the importance of key proteins required for virus propagation. This has been useful for lentivirus vector (LV) design (Ellison *et al*, 1990; Wu, 2004; Wodarz and Levy, 2007) and has allowed a focus on the use of HIV vectors as gene transfer tools (Figure 7). The first efficient LV expression system was derived from HIV-1 (Naldini *et al*, 1996a and b; Osten *et al*, 2006). Current studies suggest that the removal of HIV regulatory genes and the generation of recombinant HIV constructs will eliminate the risk of HIV pathogenesis and support the use HIV as gene transfer vectors. Kafri *et al*, 1997 demonstrated sustained gene expression in liver and muscle using HIV based vectors. Trobridge *et al*, (2008) reported the transduction of haemopoetic stem cells in pig tailed macques (Macca nemestrina) using short term *ex vivo* transduction

methods. Results showed a large number of polyclonal integrations. HIV based vectors may also be used to treat disease. Humeau *et al*, (2004) reported the stable transduction (+/-94.5%) and vector mediated inhibited HIV replication in CD4 T cells derived from HIV+ patients. An anti-HIV molecule that targets the *env* gene was delivered to CD4 T cells using a HIV based vectors. This study demonstrated the clinical application of HIV vectors in treatment of HIV and potentially delaying the onset of AIDS.

1.1.3.2 Equine infectious anemia virus (EIAV)

This non-primate virus was first described in French memoirs written by Ligne in 1843 (Leroux *et al*, 2005). Vallee and Carre demonstrated the viral etiology of persistent equine infection in 1904 by showing the causative agent was filterable (Studdert, 1996). However, EIAV exhibits unique signs of infection and associated symptoms which have not been described for other lentiviruses. This includes the occurrence of recurrent autoimmune haemolytic anaemia and fever which appears in the first year following infection. As a result EIAV was initially excluded from the RV family. However, EIAV was later confirmed as an RNA virus (Olsen, 1998). At 8.6kb the EIAV genome is the simplest and smallest of the lentivirus family (Leroux, Cadorè and Montelaro, 2004). Several studies have explored the potential use of EIAV vectors in gene therapy (Fig 6; Ikeda *et al*, 2002; Azzouz and Mazarakis, 2004; Gregory *et al*, 2004; Lamikanra *et al*, 2005). Balaggan *et al* (2006) reported the efficient and stable transduction of ocular cells following intraocular delivery of EIAV vectors by administration subretinally/intravitenally to mice. Results showed stable expression in retinal pigment epithelial (RPE) cells and photoreceptors up to 16 months post infection. McKay *et al*, (2006) also demonstrated efficient transduction of mouse tracheal cells *in vitro* and *in vivo* using EIAV vector constructs pseudotyped with influenza envelope proteins.

1.1.3.3 Feline immunodeficiency virus (FIV)

FIV (previously feline T-lymphotrophic lentivirus) infection was first diagnosed in cat sera in 1968 (Shelton *et al*, 1990). FIV was later described in 1987 soon after the AIDS epidemic (Bennet and Symth, 1992; Bendinelli *et al*, 1995; Barraza and Poeschla, 2008).

FIV infection manifests specifically in the domestic cat population (Poeschla *et al*, 1998) and cannot be transmitted to human populations. The FIV particle is 125 nm in diameter. The FIV virion is also highly glycosylated yielding surface and transmemebrane glycol proteins gp85 and gp40 respectively. Interestingly, FIV displays similar tropism, accessory proteins (Vif and RRE) and viral infection capabilities with its HIV counterpart. Similarities between HIV and FIV have made this non primate LV vector (Figure 7) ideal for gene therapy because FIV has no known pathogenesis in humans. It is hoped that FIV genotoxicity models will facilitate understanding of HIV pathogenesis and AIDS progression (Buchschacher and Wong-Staal, 2000; Sauter and Gasmi, 2000; Sauter, Gasmi and Dubensky, 2003). Stein *et al,* 2001 intravenously delivered FIV based vectors to the hepatocytes of factor VIII or β-glucoronidase deficient mice. This study reported the correction of the lysomal storage disease (mucopolysaccharidosis type VII) and hameophilliac mice and efficient transduction of transgenes into the liver of infected mice. Kang *et al*, 2005 also reported persistent factor VIII expression and partial correction of haemophilia A following the administration of FIV based vectors. Shai *et al* 2005 demonstrated the efficient and stable transduction of murine salivary glands using FIV based vectors for up to 80 days post infection. These studies all suggest the use of FIV based vectors in the treatment of disease.

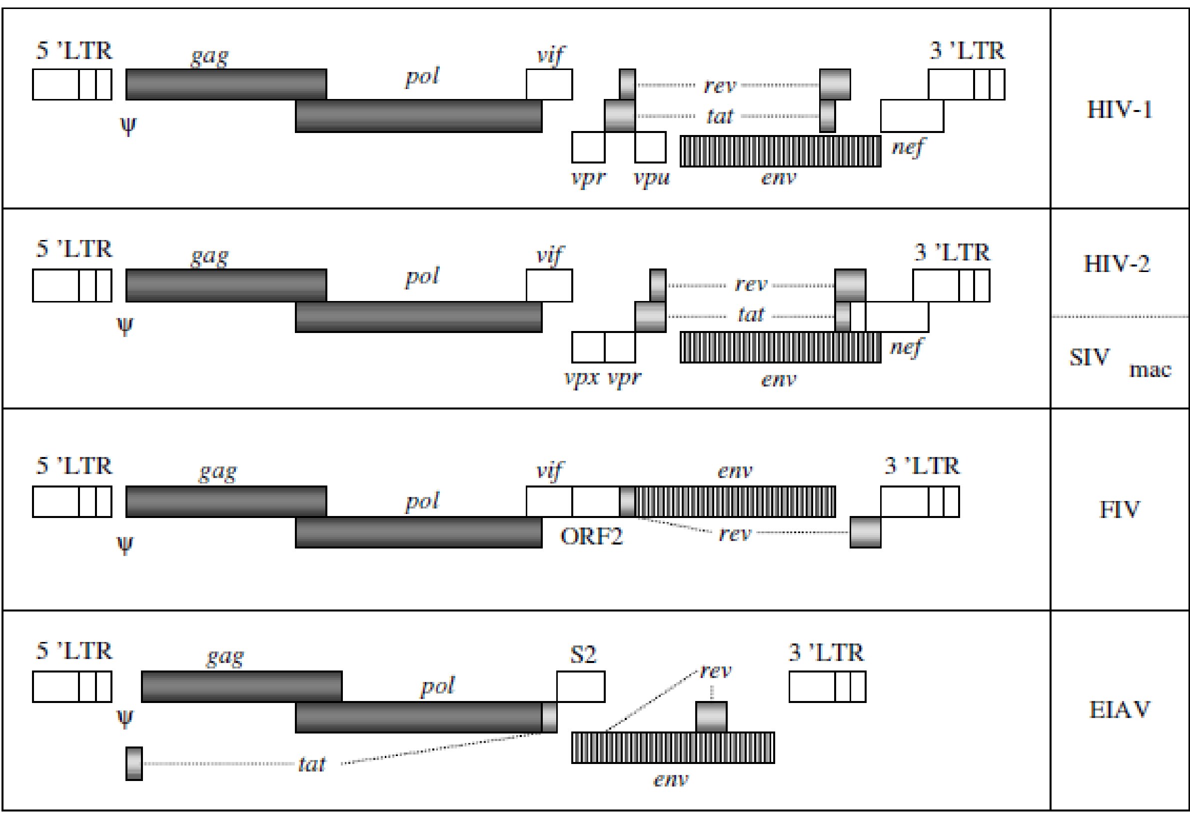


Figure 7. Genomic structure of lentiviruses that have been used in gene transfer studies. Lentiviral genome structure consists of a collection of essential structural, regulatory and accessory genes. Accessory genes are represented in white boxes and structural and regulatory genes in grey boxes respectively. Image taken from Delanda, (2004).

1.1.4 Retroviruses and lentiviruses in gene therapy

Although early gene therapy protocols relied heavily on gammaRVes these RVs are unable to penetrate the nuclear membrane of non proliferating cells such as hepatocytes, haemopoetic stem cells (HSC) and neurons (Naldini *et al*, 1996a). In contrast, lentiviruses can translocate their pre-integration complexes across the nuclear envelope independently of mitosis making them more favourable for somatic gene therapy. Naldini and coworkers (Naldini *et al*, 1996b) pioneered the effective transduction of cells both *in vitro* and *in vivo* using the HIV virus. However, lentiviruses exhibit reduced transduction efficiency of quiescent cells such as fibroblasts (Naldini, *et al*, 1996a,b). LV transduction studies facilitated the development of the first LV for clinical trials (Manilla *et al*, 2005). Recent studies have also focused on developing FIV and EIAV LV constructs for gene therapy. Problems affecting the use of LVs in gene therapy include the generation of an

immune response to the transgene product and the viral promoter driving gene expression that may be silenced. Further restrictions to retro/lentiviral integration include the action of host proteins such as Ref1 and TRIM5α since these protect mammalian cells from retroviral infection and integration by binding directly to viral capsids accelerating degradation and increasing the recognition of viral proteins as foreign agents (Saenz *et al*, 2005; Emerman, 2006; Lim *et al*, 2010; Ganser-Pornillos *et al*, 2011). By increasing vector titre the impact of host cell restriction mechanisms can be partly overcome (Barraza and Poeschla, 2008); however, this may present an increasing risk of vector related genotoxicity (Nienhuis *et al*, 2006).

1.1.4.1 The production of retroviral/lentiviral vectors for gene therapy

For most gene therapy applications it is not desirable to deliver a replication competent virus (RCV) to patients in which the vector will multiply and spread viral infection to other adjacent tissues (Chen *et al,* 1999; Hu and Pathak, 2000; Sinn *et al*, 2005). To avoid pathogenic infection viral genes (excluding *cis*-acting elements) are removed from the virus genome to form replication defective viruses (RDV) (Buchschacher and Wong-Staal, 2000). *Cis*-acting sequences are required for packaging, reverse transcription and integration steps. Helper constructs or co-transfected plasmids can be used alternatively to express viral genes removed from the virus genome in a trans-acting manner such as *env* and facilitate steps required to generate a single round of replication in target cells. Because there is a danger that recombination may occur that generates a helper virus in the transfected virus producer cells during the manufacturing process, the trans-acting viral genes have been placed on more than one plasmid construct. This circumvents the problem of a single recombination event between the transgene carrying vector and the packaging components in the producer cells generating replication competence (Chong *et al*, 1998). Therefore, ensuring whether replication competent RVes (RCR) or lentiviruses (RCL) are present is an important primary biosafety hazard that should be identified to avoid or reduce the risk of insertional mutagenesis (IM) in the host.

As a response to this potential problem split packaging systems have been developed to construct RDV's. Several helper and packaging systems exist that includes split

genome helper constructs and vector backbone plasmids that can be generated either permanently or transiently following transfection of human embryonic kidney derived 293 (Hek293) cell lines. For stable cell lines vector and helper constructs are integrated into the host cell genome of a packaging cell line (Pear *et al*, 1993; Maetzig *et al*, 2011). This usually results in the production of high amounts of retroviral proteins required for assembling virus particles to achieve high titre (Pear *et al*, 1993; Ory, Neueboren and Mulligan *et al*, 1996; Poeschla *et al*, 1998). Production of virus particles by stable producer cell lines is time consuming and viral components such as the VSV-G envelope may not be tolerated since they are cytotoxic. Moreover, the efficiency of stably transfected cell lines depends on the long term survival of successfully transfected cells and isolation of high titre producer clones from populations of mixed titre producing cells (Brielmeier *et al*, 1998). The transient transfection method is an alternative and is flexible and rapid as virus particles are harvested within a few days after plasmids transfection of 293T. This cell line is widely used in transient transfection systems to express recombinant proteins since nearly 90% transfection levels can be achieved (Thomas and Smart, 2005).

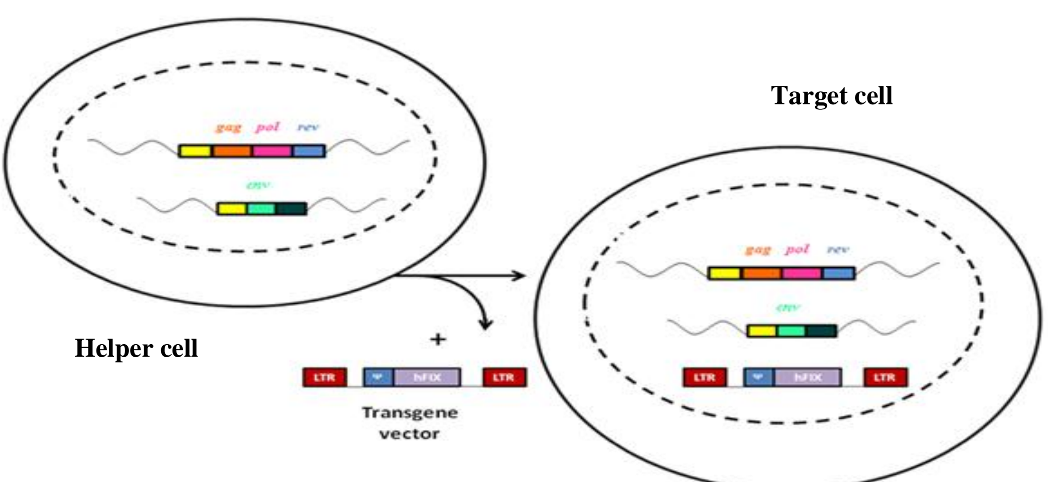


Figure 8. Production of RV vectors for gene therapy. The viral proteins required for transgene expression are separated into several packaging plasmids to reduce the risk of generating replication competent viruses. Hence, the *gag* and *pol* genes are included in one construct, the *env* gene is included in another plasmid and the vector 'backbone' carrying the transgene are on a third plasmid. Wild type envelope sequences can be replaced with alternative *env* genes in order to alter viral tropism in specific cells or tissues. Expression of the therapeutic gene is driven by the strong promoter signals derived from the LTR. The complete retroviral construct is assembled and released from the producer cell for use in *in vitro* and *in vivo* studies.

Self Inactivating (SIN) vectors have been designed and used in gene transfer methods to reduce the risk of generating replication competent viruses during the production of viral vectors (Miyoshi *et al*, 1998). SIN vectors contain a deletion of promoter and enhancer sequences present in the U3 region of the LTR (Yu *et al*, 1986; Miyoshi *et al*, 1998; Suerth *et al*, 2010). This prevents the production of vector length (genomic) transcripts, replication of virus particles and promoter activation of host genes adjacent to provirus inserts sites (Olson *et al*, 1994).

Dull *et al* (1998) first reported high titre production of HIV based vectors containing a SIN LTR configuration. CMV sequences were used to replace upstream promoter sequences and create a 5LTR chimeric transfer construct for the transduction of rat brain cells (Dull *et al*, 1998). Most importantly, SIN vectors are still capable of achieving a high level gene transfer and stable gene expression in infected cells (Miyoshi *et al*, 1998).

Figure 9. The sites of sequence deletions are represented by dashed arrows. The removal of transcriptional elements is used to generate SIN LTR constructs. Modified U3 sequences are indicated by Δ. Image taken from Suerth *et al*, (2010).

1.1.4.2 Successful gene therapy clinical trials using viral vectors a history

In 1990, the first approved human gene therapy trial was performed for the treatment of adenosine deaminase (ADA) deficiency by Anderson and co-workers. The ADA enzyme is essential for catalysing the deamination of adenosine and the ADA substrate 2′-deoxyadenosine (dAdo) into inosine deoxyinosine, respectively (Muul *et al*, 2003). Mutations of the ADA gene results in enzyme deficiency, the accumulation of dAdo and lymphoid toxicity (Joachims *et al*, 2008). Furthermore ADA-deficient mice exhibit severe reductions in T, B and natural killer (NK) cell

populations which if untreated results in severe combined immunodeficiency disease (SCID) (Blackburn *et al*, 1996; 2000). Blaese and Anderson (1995) successfully achieved retroviral mediated transfer of the ADA gene into T lymphocytes *ex vivo*. Transformed lymphocytes were then returned to the bone marrow of two recipient children restoring T cell populations and ADA gene expression (Kohn *et al*, 1989; Culver *et al*, 1991; Blaese *et al*, 1995). A follow up study by Muul *et al*, (2003) reported the persistent expression of ADA genes driven by first generation retroviral vectors 12 years after treatment. This result supports the long lasting effects of retroviral mediated gene transfer. Furthermore no adverse effects were reported in both patients. However, both patients continued to receive a reduced dose of enzyme replacement treatment with polyethylene glycol-conjugated bovine ADA (PEG-ADA) which demonstrates that only partial correction of metabolic abnormalities had occurred.

Unfortunately, in 1999 gene therapy saw its first setback with the death of Jesse Gelsinger after treatment for ornithine transcarbamylase deficiency (OTC) using first generation adenovirus (Smith and Byers, 2002; Thacker, Timares and Matthews, 2009). In 2000, however, gene therapy re-emerged as a plausible therapeutic technique following successful clinical treatment of X-linked severe combined immuno-deficiency (X-SCID) disorder (Cavazzana-Calvo *et al*, 2001). This trial, conducted in France by Marina Cavazzana-Calvo and Alain Fischer demonstrated the ability to correct the disease phenotype by *ex-vivo* gene transfer to patient cells using a Moloney RV derived vector (Cavazzana-Calvo *et al*, 2000). Other promising developments in gene therapy have been observed for the treatment of X-SCID (Gaspar *et al*, 2004; Fischer, Hacein-Bey-Abina and Cavazzana-Calvo, 2010; Baum *et al*, 2011; Gaspar *et al*, 2011; Kennedy *et al*, 2011), ADA SCID (Aiuti *et al*, 2002a,b; 2009) and individuals suffering from chronic granulomatous disorder (CGD) (Malech *et al*, 1997; Seger, 2008; 2010) cancer (Ciceri *et al*, 2005; Morgan *et al*, 2006), haemophillia (Lu *et al*, 1993; Mátrai *et al*, 2010) congenital blindness (Hauswirth *et al*, 2008; Smith *et al*, 2009; Bainbridge, 2009) muscular dystrophy (Wells, 2006; Brolin and Shiraishi, 2011; Pichavant *et al*, 2011) and viral infections (Von Laer *et al*, 2009). Years after the successful correction of X-SCID gene therapy underwent a second setback with the development of leukaemias in 5 (4 in French trial and 1 in British trial) out of 19 patients treated for X-SCID (Davé et al,

2009; Qasim *et al*, 2009). This provoked several studies that aimed to characterise the factors that may be involved in oncogenesis.

1.1.5 Retroviruses and genotoxicity in the host

Viral integration can occur in a large selection of genes with various functions. Several studies have shown that retroviral integration is a semi-random event and may contribute to IM (Mitchell *et al*, 2004; Deichmann *et al*, 2007).

Insertion of DNA fragments is an important step in retroviral infection and occurs naturally during transposition of mobile elements; however, retroviral integration can disrupt host genomic sequences (Singhal *et al*, 2011).

In addition to earlier genotoxicity studies (Mirsalis *et al*, 1995; Recio *et al*, 1995) models have been developed that assess genotoxicity with a focus on identifying genes such as proto-oncogogenes that may be associated with IM. From these studies it has become apparent that viral integration and the risk of IM are influenced by target site selection. Target site selection is crucial to virus insertion and influences host responses following infection (Holman and Coffin, 2005). Therefore, understanding integration patterns is central to the future clinical use of viral vectors in human gene therapy trials. RV integration is not random nor sequence specific as previously demonstrated (Wu *et al*, 2003; Shröder *et al*, 2002; Ciuffi *et al*, 2005). This is despite the identification of weak pallindromic sequences found associated with HIV-1, EIAV and RV integration (Carteau *et al*, 1998; Wu and Burgess, 2004; Holman and Coffin, 2005; Hacker *et al*, 2006).

Shröder *et al* (2002) analysed over 500 integration events of HIV-1 and RV vectors following infection of a human T cell line. Mitchell *et al* (2004) investigated 3,127 integration sites of ASLV, MLV and HIV-1 infection and together with several other integration studies also involving ASLV, oncoRVes and lentiviruses unique integration patterns have been revealed (Bushman, 2003; Wu *et al*, 2003).

Gamma RVes such as MLV prefer to integrate into the vicinity of transcription start sites, promoter regions and CpG islands which facilitate gene regulation (Wu *et al*, 2003; Bushman *et al*, 2005; Lewinski *et al*, 2006). This greatly increases the potential for RV mediated IM. These integration patterns have been supported by

cross species integration patterns identified in Human and Rhesus-Macaque infected cells (Bushman *et al*, 2005).

In contrast, lentiviruses such as HIV and EIAV prefer integration into gene transcription units which is believed to be safer for gene therapy (Shröder *et al*, 2002; Bushman *et al*, 2005; Hacker *et al*, 2006; Cherepanov, 2007). The insertion profile for EIAV vectors, described by Hacker *et al* (2006) showed that these viruses favour integration into gene dense and highly expressing genes and therefore it may be suspected these vectors could be more likely to disrupt gene expression than HIV based vectors. In comparison ASLV viruses exhibit weak preferences for integration into transcriptional units and appear to randomly integrate into the host genome.

Studies conducted by Lewinski *et al* (2006) implicated the role of viral sequences such as those of the integrase and *gag* genes being involved in target site selection. However, tethering interactions between viral proteins and cellular proteins may also influence viral target site selection. Lens epithelium-derived growth factor (LEDGF) is a nuclear chromatin associated protein that binds to integrase (Cherepanov, 2007). It has been suggested that LEDGF acts as a tethering factor for integration of lentiviruses although this is only known for HIV (Ciuffi and Bushman, 2006). HIV infected cells depleted of LEDGF1/p75 protein (by RNA interference) exhibit a lower frequency of integration into transcription units when compared to cells containing these proteins (Lewinski *et al*, 2006; Kang *et al*, 2006; Ciuffi *et al*, 2006). The viral enzyme *integrase* has also been shown to play a direct role in tethering interactions and target site selection. Lewinski *et al*, (2006) found that transferring the MLV *integrase* into HIV vector constructs caused the hybrid to develop MLV integration patterns. In addition the inclusion of both MLV gag and *integrase* genes into HIV LV constructs further increased the similarity of target site selection to that of RV vectors.

Integration into centromeric heterochromatin has also been shown to be unfavourable for HIV integration (Bushman *et al*, 2005) suggesting that the transcriptional activity of target cells plays a crucial role in integration target site selection (Alabanese *et al*, 2008).

Recently Biasco *et al*, (2011) investigated the effects of MLV integrations on clonal expansion and the frequency of common insertion sites (CISs) in haematopoietic cells following gene therapy. This study was conducted in five ADA-SCID patients

who received haemopoetic stem cell gene therapy without any observable adverse effects. The insertion profile of MLV was analysed in both peripheral blood and HSC samples taken from patients before and several years after treatment. This study revealed cell specific insertion patterns in peripheral blood lymphocytes (PBL) when compared to haemopoetic stem cells (HCS's). Increased numbers of integrations into genes involved in the immune systems and DNase hypersensitive sites were favoured in PBL's when compared to HSC's. IM of host candidate cancer genes is believed to play a key role in uncontrolled cellular proliferation and malignant growth. However, other cellular intrinsic factors may contribute to oncogenesis following viral integration. Epigenetic regulation occurs via mechanisms such as methylation that control gene expression. It is possible that lentivirus integration represents a foreign agent that elicits an epigenetic response and induces uncontrolled cellular proliferation. Biasco *et al*, (2011) demonstrated that vector insertions are cell specific but an increasing number of insertions were observed in transcriptionally active chromatin sites marked by histone modifications. This supports the possibility that virus integration may elicit an epigenetic response. Another factor which may influence target site selection is DNA distortions. DNA bending proteins such as nucleosomes may distort DNA targets prior to integration and this activity can promote integration (Shröder *et al*, 2002).

1.1.5.1 Genotoxicity following RV and LV infection of the host genome

Previously Soriano, Gridley and Jaenisch (1987) showed that retroviral insertion into a specific locus can induce a recessive lethal mutation resulting in embryonic death. IM is used to describe the modification of target DNA via the incorporation of additional bases (Singhal *et al*, 2011). IM often occurs due to virus integration into regulatory and coding sequences required for 'normal' gene expression. IM may also disrupt target DNA by inducing chromosomal rearrangements and recombination that promote chromosome instability (Bardwell 1989; Themis *et al*, 2003; Baum *et al*, 2004; Uren *et al*, 2005; Suzuki *et al*, 2006). Hence, several mechanisms of IM exist. Before discussing the risks of virus-mediated IM in gene therapy it is important to describe the currently views on mechanisms that appear associated with IM.

1.1.5.2 Retroviruses and aberrant splicing

Integration by proviruses (Figure 10 a-b) into gene coding sequences may ultimately generate proteins with oncogenenic potential due to deletion of either the N terminal (NH2-) or C terminal (COOH) domains (Uren *et al,* 2005). Jiang *et al* (1997) reported the truncation of the proto-oncogene *c-myb* following proviral integration of a recombinant ALV vector. In this case the truncated *Myb* protein was found to be highly oncogenic due to deletion of the N terminal domain which resulted in the induction of several tumours and B cell lymphomas. Truncation may be caused by aberrant splicing or 3' promoter activation after insertion within a gene or genome rearrangements involving viral sequences (Ceci *et al*, 1997; Jiang *et al*, 1997; Uren *et al*, 2005).

1.1.5.3 Retroviruses and promoter activation

Temin *et al* (1993) previously described the importance of retroviral LTR sequences for integration and termination of transcription and polyadenylation of RNA. Integration of retroviral sequences can result in transcriptional activation of neighbouring genes via retroviral promoters and enhancers (Figure 10c). The RV must integrate in the same transcriptional orientation as the neighboring host genes before activation of transcription by promoter sequences can occur (Mölder *et al*, 1985). LTR insertion can result in proto-oncogene activation from either 5' or 3' LTR promoter. Herman and Coffin (1987) transduced chicken embryo fibroblasts with RVes and found that the promoter activity of the 5LTR was 50 times higher than that of the 3LTR in wild type ALV RVes. However, the 3' LTR promoter is thought to play a key role in proto-oncogene activation. Interestingly, a separate study demonstrated that ALV can activate *c-myc* by transcriptional activation from the 3' LTR following provirus integration upstream of the c-myc coding sequences (Boerkoel and Kung, 1992). This reduced transcription of the 5' LTR is thought to be due to deletions located close to this LTR.

1.1.5.4 Retroviruses and enhancer activation

Activation of transcription by U3 enhancer sequences has a bidirectional effect on flanking DNA sequences regardless of the orientation of the integrated virus (Mölder *et al*, 1985).

Recent studies have demonstrated that a single insertion in combination with transgene enhancers may influence oncogene expression up to 100kb from the provirus insertion site (Hanlon *et al*, 2003; Baum *et al*, 2004; Nienhuis, Dunbar and Sorrentino, 2006).

1.1.5.5 Retroviruses and read through transcription

The expression of structurally aberrant proteins can result from chimeric transcripts that are initiated within the RV LTR sequence and read-through into downstream cellular sequences (Figure 10d; Mölder *et al*, 1985; Conley *et al*, 2008). Read-through transcripts can be packaged into virions as effectively as normal viral RNA. These mature chimeric transcripts can result in the increase of viral and flanking cellular genes (Herman and Coffin, 1986; Nienhuis, Dunbar and Sorrentino, 2006). Singhal *et al* (2011) have concluded that the removal of the viral polyadenylation signal present in the LTR could prevent the occurrence of hybrid RNA generated by read-through transcription following integration into intron sequences.

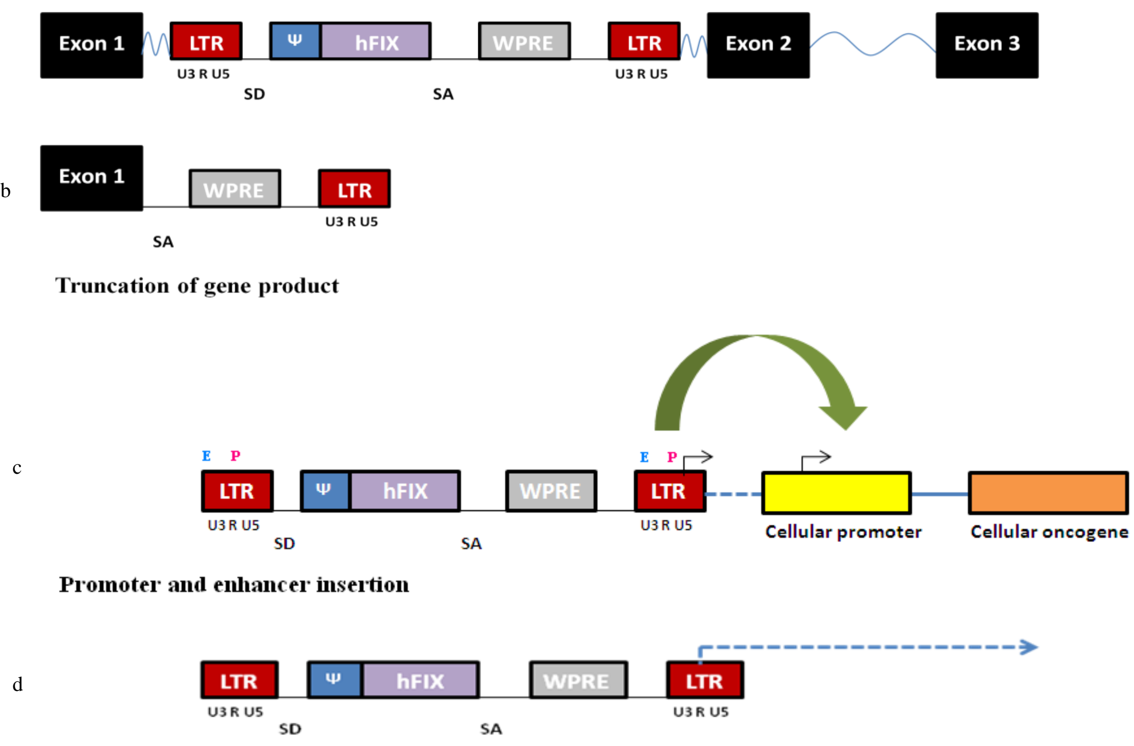


Figure 10. Mechanisms of insertional mutagenesis (IM) associated with retroviral integration. a, b) Truncation of gene product. Vector insertion can disrupt normal splicing mechanisms yielding aberrant gene products. This may result in the inactivation or mutation of protein sequences. c) Promoter insertion. Promoter sequences present in the LTR region can result in the activation of adjacent cellular promoters and oncogenes. Enhancer and promoter sites located within the LTR are annotated in blue and pink respectively. d) Read-through transcription. LTR can drive the production of transcripts that read through from proviral sequences into downstream cellular genes.
LTR: Long terminal repeat; Ψ: Packaging signal; hFIX: human factor IX (transgene); WPRE: Woodchuck hepatitis post transcription regulatory element; SD: Splice donor; SA: Splice acceptor.

1.1.6 Retrovirus genotoxicity in gene therapy trials

1.1.6.1 X – Linked SCID trials

The γ-chain is a key component of several cytokines including IL-2 and is important for the development of T and natural killer cells (Hacein-Bey-Abina *et al*, 2002; Howe *et al*, 2008; Ginn *et al*, 2010). X – Linked SCID is caused by a mutation in the gene encoding the IL2-receptor γ-chain (IL2RG). Disruptions in IL-2 receptor signalling results in the dramatic reduction of T-cells, natural killer cell populations

and non-functional B cells. Reports show that X-Linked SCID accounts for approximately 40-50% of all SCID cases (Gaspar *et al*, 2004; Howe *et al*, 2008). This X-linked disease is particularly obvious in the first year of life due the occurrence of recurrent infections (Hacein-Bey-Abina *et al*, 2003). Successful treatment of this disorder relies on allogenic bone marrow transplantation from HLA matched donors to deliver healthy γC genes. This is often followed by administration of immunoglobulins (Cavazzana-Calvo *et al*, 2005; Thrasher *et al*, 2006). However for most patients well-matched donors are not available and bone marrow transplantation from mis-matched donors may result in death. To overcome this problem research to treat and possibly correct X-SCID has been attempted using gene therapy of autologous haemopoetic stem cells. Pre-clinical models of haemopoetic gene therapy showed promise for the treatment of X-SCID. Restoration of T, B and NK cells was achieved in γC deficient mice following retroviral mediated delivery of human γC genes (Lo *et al*, 1999). However, preclinical studies used to assess the safety of gene therapy vectors were slow to predict the risk of gene transfer. The failure to predict the likelihood of IM and associated oncogenesis was due to three major factors. Short follow up studies, analysis of insufficient number of insertion sites and subjective quantitative assessment of genotoxic risk (Porteus, Connelly and Preutt, 2006). Human clinical trials on gene therapy for X-SCID took place in Paris (Cavazzana-Calvo *et al*, 2000; Hacein-Bey-Abina *et al*, 2002) and London (Gaspar *et al*, 2004) and showed sustained correction of this disease in several patients. The trials involved transduction of autologous CD34+ bone marrow cells with gammaretroviral vectors containing the γC gene that were transduced *ex vivo* and returned to each patient. Antibody levels and immune cell populations were measured following treatment and both studies reported restoration of cellular and humoral immunity months after treatment (Qasim, Gaspar and Thrasher, 2009; Hacein-Bey-Abina *et al*, 2010). Unfortunately, 5 patients that received gene modified bone marrow cells developed leukaemia 3-5 years after treatment (Hacein-Bey-Abina *et al*, 2010). RV insertion sites were retrieved using a highly sensitive PCR assay known as linear amplification mediated (LAM) polymerase chain reaction (PCR). This technique combined with sequencing analysis has been used to find individual insertions into host genomic sequences by the provirus. A number of techniques are now used

routinely for the detection and profiling of proviral insertion sites that include inverse PCR and ligation mediated (LM - PCR) and modifications to these such as non-restriction (nr) LAM PCR coupled with 454 pyrosequencing. These allow rapid identification of thousands of proviral insertions to characterise insertion site preferences and hence the potential for a chosen integrative viral vector to cause IM.

Using X-SCID patient material follow up studies and integration site analysis revealed that RV integration had occurred into proto-oncogenes and genes implicated in leukamogenesis. Out of the 5 patients, 3 patients in Paris and 1 in the London trial exhibited insertional activation of the proto-oncogene LIM domain only 2 (LMO2) (Dave *et al*, 2009; Qasim, Gaspar and Thrasher, 2009; Hacein-Bey-Abina *et al*, 2010). Treatment of large numbers of cells in X-SCID gene therapy is essential to achieve efficient immune reconstitution. However, high transgene expression is believed to be involved in IM (Kohn *et al*, 2003; Baum *et al*, 2004). The over-expression of *IL-2YRG* gene can cooperate with LMO2 and induce tumour formation following the delivery to X-SCID patients (Dave *et al*, 2004; 2009; Woods *et al*, 2006). In a separate study, Modlich *et al* (2005) reported that the expression of the multidrug resistance gene 1 (MDR1) may be associated with the development of leukaemias in mice following retroviral gene transfer. In these studies murine bone marrow cells were transduced with retroviral vectors carrying the MDR1 transgene. Unfortunately, transduced cells had high transgene copy number and multiple insertions were found in proto-oncogenes in malignant clones. Although several studies have reported successful transduction of bone marrow cells with MDR1 this study revealed several issues. The occurrence of leukaemia may have been driven by increased MDR1 expression or genomic instability associated with high vector copy number (Modlich *et al*, 2005).

X-SCID mouse genotoxicity models have also been developed to assess the cause and risk of vector mediated IM that occurred in the X-SCID infants (Qasim, Gaspar and Thrasher, 2009). Using the most currently developed models the integration profiles of lentiviral vectors such as HIV into the gene transcription unit rather than promoter regions suggest lentiviruses have a reduced risk of IM. A study by Ginn *et al*, (2010) investigated the efficiency and safety of SIN lentiviral vectors used to drive expression of the human γC transgene in X-SCID mice. Interestingly, results showed lymphoma development in 4 out of 14 mice treated with lentiviral vectors.

However, this study failed to implicate IM or γC over expression as the cause lymphomagenesis. Several studies that demonstrated vector induced oncogenesis have also reported transgenic expression of growth promoting signalling molecules (Baum *et al*, 2003). These studies showed that signalling alterations caused by the transgene and IM may cooperate to induce activate oncogenes. A mouse study conducted by Li *et al* (2003) introduced a truncated version of the p75 neutrotrophin receptor (p75NTR) known as dLNGFR. Both p75NTR and the artificial mutant – dLNGFR may associate with the tyrosine kinase receptors (trk) in neutrophins. However, the cytoplamic domain located in the p75NTR has important pro-apoptotic functions. The retroviral vector inserted and upregulated the proto-oncogene Evi1. Li *et al* (2003) found that interaction between dLNGFR, neurotrophins and trkA receptors in Acute Myeloid Leukaemia (AML) cells contributed to the transformation of leukemic cells and development of an unusual form of monocytic leukaemia. This supports the importance of dysfunctional signalling in oncogenesis. A previous study by Hantzopoulos *et al,* (1994) also showed that interaction between dLNGFR, nueurotrophins and trk receptors failed to signal apoptotic/differentiation pathways required for balanced cell growth. As a result dLNGFR was able transform normal fibroblasts. Meyer *et al*, (2007) demonstrated the ability of mutant TrkA neurotrophin receptors to induce AML. These studies all highlight the role of signal interference induced by the transgene product dLNGFR in mutagenesis. These studies also describe the importance of cooperation between vector, transgene and host signalling molecules. Deichmann *et al*, 2007 examined retroviral integration sites in CD34+ transduced cells and peripheral blood cell samples obtained from X-SCID patients. These patients were recipients in the first X-SCID gene therapy trial. Their study found that two thirds of insertions occurred commonly in or near to genes involved in cell signalling growth and the cell cycle. Other retroviral integration studies have revealed unique integration patterns that have occurred as a result of influences by the vector or the host that are, as yet, unclear. This has, therefore triggered debate on the choice of vectors to be used in gene transfer.

1.1.6.2 Chronic Granulomatous Disease (CGD) trial

X-Linked CGD is a recessive inherited immunodeficiency caused by the absence of NADPH oxidase activity. NADPH oxidase is an enzyme that catalyses the production of superoxide from oxygen and NADPH (Babior, 2004). NADPH oxidase is composed of several genes including *p22phox* (*CYBA*); *P67phox* (*NCF2*) and *gp91phox* (*CYBB*). Mutations in these genes result in CGD however, two thirds of CGDs are caused by mutations in the *gp91phox* gene (Björgvinsdóttir *et al*, 1996; Dinauer *et al*, 1996, 1999; Saulnier *et al*, 2000; Brenner *et al*, 2006). In the absence of NADPH oxidase activity neutrophils and other phagocytic leukocytes are unable to generate superoxides required to kill microbes. As a result CGD is characterised by recurrent severe bacterial and fungal infections (Stein *et al*, 2010). RV-mediated gene transfer of the *gp91phox* gene into bone marrow cells has been conducted by several groups.

Successful transduction of murine X-CGD bone marrow cells has been reported by several groups. Transduced bone marrow cells containing *gp91phox* genes were used to treat X-CGD and NOD/SCID mice (Björgvinsdóttir *et al*, 1996; Dinauer *et al*, 1999; Brenner and Malech, 2003; Sadat *et al*, 2009). These studies reported superoxide production, correction of phagocytes and improved phagocyte function. RV mediated delivery of *gp91phox* genes has also been reported in male rhesus monkeys (Brenner *et al*, 2006). In this study CD34+ cells were isolated from healthy rhesus monkeys (*Macaca mulatta*) and cells were transduced with an MLV vector expressing the *gp91hox* gene. Analysis of RV of insertions sites revealed integration in the vicinity of genes involved in cell cycle and proliferation. However, the use of γ-RVes in all of these studies failed to generate the clonal imbalance found in the X-SCID trials.

Unfortunately, a similar trial conducted in human patients resulted in clonal expansion, myelodyplasia and monosomy of chromosome 7 and death of a patient as a result of IM into the MDS1-Evi1 locus. Autologous haemopoetic stem cells had been transduced and returned to the patient. Increased gene expression in these dominant clones was linked to genome instability and raises an important concern that the cause of genotoxicity associated with integration into the host genome is still unclear and whether *in vivo* clonal selection contributes to clonal dominance or not

(Cattogolio *et al*, 2007). In the X-SCID, since out of the 5 patients who developed leukemia 4 showed provirus insertion and over-expression of the *LMO2* gene it is believed that cooperation between the transgene (*IL2rg*) and *LMO2* could have contributed to leukomogenesis. Whatever the view, these observations highlight the risk of cancer development following gene therapy using retroviral vectors.

Most recently, clonal dominance has also been identified in 2 adult patients treated for CGD 15 and 28 months after treatment. In both CGD patients the retroviral vector used for gene transfer has been shown to insert into the MDS1-Evi1 locus. Specific clones containing the MDS-Evi1 insertions dominated the CD34+ haemopoetic stem and progenitor cells resulting in the over expression of the *Evi1* gene. The expansion of clones containing MDS-Evi1 insertion sites was also seen *in vitro* in bone marrow cells and a rhesus monkey treated with MLV (Calmels *et al*, 2005; Du and Copeland, 2005; Ott *et al*, 2006; Stein *et al*, 2010). Dunbar and co workers (2010) suggest that the up-regulation of the Evi1 gene may promote chromosome instability, chromosome abnormalities and disease progression (Dunbar and Larochelle, 2010). Interestingly, other studies have reported the absence of recurrent MDS1-Evi1 insertions and leukaemias despite *ex vivo* gammaretroviral gene transfer into haemopoetic stem cells (Kuramoto *et al*, 2004a, b; Brenner *et al*, 2006). This suggests that additional factors may be involved in the selection of insertional mutants.

Taken together, the findings of the human X-SCID and CGD clinical trials emphasise the importance of genotoxicity and IM in gene therapy resulting from integrating viral vectors. Interestingly, Soudais *et al*, (2000) demonstrated when treating X-SCID and recovery of immune function in γc and JAK3-deficient mice with RVes carrying the γc gene no adverse effects were found even when alternative gene transfer methods were used. When Aviles Mendoza, (2001) also demonstrated immune reconstitution in X-SCID deficient mice using five retroviral vectors expressing the *IL2RG* gene no prediction could be made on the risk of IM following treatment for SCID – X1 and therefore it was doubtful that pre-clinical models can predict vector safety accurately. Hence, although these models are important to satisfy regulatory bodies such as the US Food and Drug (FDA) and European medicines agency (EMA) no model as yet satisfies fully our understanding of the likelihood and causes of genotoxicity.

1.1.7 Models for retrovirus genotoxicity

Genotoxic assays have been previously used to assess the mutagenic risk of agents including chemicals and retroviral vectors (Mirsalis *et al*, 1995a, b; Recio *et al*, 1995; Okada *et al*, 1997; Reliene *et al*, 2003; Modlich *et al*, 2006). The mouse is the leading mammalian model used in genotoxicty research. They are preferred tools because of their rapid breeding times and the extensive knowledge of their genetic makeup. Larger animals including rhesus monkeys have also been useful as models to assess gene therapy safety (Tarrantal *et al*, 2001). However, other organisms such as the *Drosophila* fly (Cooley, Kelley and Spradling, 1988), rhesus monkeys (Seggewiss *et al*, 2007), Sheep, Pigs, Dogs and Zebrafish (Amsterdam *et al*, 1999; Haffter *et al*. 1996) and *in vitro* studies have been used to assess the mutagenic risk of retroviral vectors.

The *hprt* assay has been widely used in research applications to study the effects of single gene mutations on human and rodent cells. The *hprt* gene present on the X-chromosome encodes the hypoxanthine-guanine phosphoribosyltransferase (HPRT) enzyme which is required in the purine salvage pathway. The HPRT enzyme catalyses the transformation of purines to monophosphates which are cytotoxic to normal cells. However, the loss of *hprt* activity and tolerance to treatment with harmful purine analogues can be induced by mutations in the *hprt* gene. King and Brookes (1985) described the *hprt* gene as a useful target for carcinogenic and mutagenic agents. King and Brookes (1985) also reported the use of Chinese Hamster V79 cells which contain a single copy of the *hprt* gene to study HPRT mutagenesis. This study found that V79 cells treated with the carcinogen MCDE induced *hprt* mutants and resistance to the purine analogue 6-thioguanine (King and Brookes, 1985) Hence, cells carrying *hprt* mutations can be easily distinguished from unmutated cells in culture. Tates (1991) described the use of a cloning assay to select *hprt* mutants in T-lympocytes derived from patient samples. T lymphocytes were collected from healthy subjects, cancer patients and subjects exposed to the carcinogen ethylene oxide (Tates 1991). This study revealed a significant increase in mutation frequencies in T-cells containing *hprt* mutants. This supports the role of carcinogens in mutagenesis. In 1995 Lichtenauer-Kaligis *et al* investigated the influence of genomic position on integrated *hprt* cDNA in the human genome. The

hprt cDNA gene was integrated into human lymphoblastoid TK6 cells. Results showed a spectrum of 100 mutations in the *hprt* gene (Lichtenauer-Kaligis *et al*, 1995). Other studies have explored the risk of IM by replication defective RV's in mammalian cells. King *et al* (1985) found that MoMLV was able to inactivate the *hprt* gene in embryonic carcinoma cells. Goff predicted the frequency of mutations induced by a single provirus insertion at a single locus. The estimated frequency given was one inactivating mutation in 10^{-6} virally exposed cells (Goff, 1987). Themis *et al,* (2003) demonstrated the ability of RV integration to inactivate the *hprt* gene in haploid V79 Chinese hamster cells. This study confirmed that an attenuated defective RV virus was able to induce mutations at a single gene locus. Themis *et al,* (2003) also reported a 2.3 fold increase in the risk of mutagenesis where multiple provirus insertions had occurred into the host cell genome and implicate the role of vector copy number in insertional mutagenesis. Results showed that the occurrence of mutagenesis correlated with high titre and multiplicity of infection (MOI) also known as the number of transducing viral particles per cell.

This supports the notion that virus MOI should be reduced to avoid mutagenesis. Kustikova *et al*, (2003) also investigated the relationship between vector copy numbers and gene transfer efficiency in primary CD34+ haemopoetic, peripheral blood and K562 leukaemia cells. This *in vitro* study revealed a linear correlation between vector copies and gene transfer efficiency. An increase in vector integration and copy number resulted in higher transduction rates. A single vector insertion into a cell resulted in a gene transfer rate of less than 30%. However, the occurrence of 9 vector insertions increased the transduction efficiency to approximately 90%. This study also suggested the occurrence of IM may correlate with the number of integrated vector (Kustikova *et al*, 2003) and thus multiple vector insertions can promote genotoxic effects such as malignant growth. Early studies conducted by researchers including Stocking *et al* (1993) were responsible for the predicting low (10^{-6}-10^{-8}) theoretical risk of viral vector mediated mutagenesis (Copeland, Hutchison and Jenkins, 1983; Moolten and Cupples, 1992).

Fehse *et al* (2004) developed a statistical model based on Poisson's mathematical formulae which could be applied to mutagenesis studies. This model aimed to estimate the number of retroviral insertions and assessed the distribution of integrated vector copies in a population of cultured cells. This model provided a

useful tool for predicting vector copies in cell populations and also found that multiple vector insertions increased the risk of IM and oncogenesis. The human genome has been reported to contain more than 20, 000 genes (Frazer *et al*, 2004; Stein, 2004; Barry *et al*, 2007) which may be vulnerable to retroviral integration and mutagenesis and hence, increase the potential risk for proto-oncogene activation.

Modlich *et al* (2005) investigated the effects of vector dose on retroviral transduction. RV vectors containing the MDR1 gene were used to transduce bone marrow cells *ex vivo* then returned to C57BL/6J mice (Modlich *et al,* 2005). This study again demonstrated that increased vector dose contributed to IM (Modlich *et al,* 2005). Clones containing multiple RV insertions into (or in the vicinity) proto-oncogenes and signaling genes were over-represented in this study suggesting that leukamic growth arose due to the expansion of dominant clones and hence clonal imbalance triggered by vector insertion may initiate malignant transformation. Modlich *et al*, (2006) developed a cell culture assay using primary bone marrow cells to study the effect of vector dose on IM. Isolation of retroviral insertions revealed a correlation between MOI and the incidence of immortalised clones caused by insertions into the *Evi1* gene. Montini *et al*, (2009) used a tumour prone Cdk2na−/− knockout model to predict the genotoxic risk of viral vectors used for clinical studies. The integration of RV and LV vectors into cancer causing genes induced tumour development. This study showed that LV vectors carrying active LTR's were 10 times safer than their RV counterparts and this was thought due to the integration site selection of RV in target cells. Kustikova *et al* (2007) used long term murine studies to generate a database of RV insertion sites that had caused malignant clonal dominance. To do this bone marrow stem cells were cultivated *ex vivo* and delivered to mice. This study suggested that specific RV hits into genes that regulate proliferation, apoptosis, signaling and transcription contributed to the occurrence of clonal expansion. Researchers including Stocking and co-workers (1993) contributed greatly to the development of *in vitro* models for analysis of retroviral IM. Stocking *et al* (1993) developed a tissue culture system using a human myeloid stem cell line (TF-1). Using retroviral vectors to induce mutations a direct correlation between retroviral insertions and a mutation frequency of 2.2×10^{-7} was demonstrated.

It has been suggested that transformation of cell populations with large cell turnover such as haemopoetic stem cells may contribute to clonal dominance and expansion of clones containing harmful retroviral insertions. Kustikova *et al* (2009) investigated the impact of cell sorting and purification of haemopoetic stem cells on IM. HSC's were subdivided into 2 groups primitive HSC's (LSK) and purified HPC's (LK). This study found that improved purification of stem cells did not reduce the genotoxic effects of γ-retroviral transduction. However, limiting the number of HSC's transduced did reduce the genotoxic risk of retroviral gene transfer. In the same study Kustikova *et al* (2009) used lentiviruses to transduce a restricted number of purified HSC's. This reduced the risk of clonal imbalance caused by provirus insertions into proto-oncogenes despite the high transduction efficiency of these vectors when compared to γRVes. These findings further implicate the role of target cell type, the number of transduced cells and vector type as variables to consider when assessing the risk of gene therapy.
Genotoxicity models are useful tools for uncovering the transforming ability of viral vectors and viral integrations which induce cancer development. Large scale analysis of vector integration will continue to rely upon previous and current genotoxicity models to evaluate the risks and benefits of using RV vector in gene therapy.

1.1.7.1 Fetal gene therapy model

The concept of fetal (*in utero*) gene therapy is based on the prevention of fatal genetic conditions and achieving permanent correction of such diseases by stable transduction of cell populations before they manifest in the unborn child (Coutelle *et al,* 2005).
The application of somatic gene transfer to the foetus is relatively new to the field of gene therapy. The efficiency of somatic gene transfer into adult cells is limited by various factors including the failure to find MHC compatible donors and the development of an immune response mounted against transgene sequences following administration of viral vectors. Adult somatic gene therapy is also limited by the large target cell to vector ratio. Fortunately, research has shown that these problems could be avoided by treating monogenetic disorders *in utero*.

The developing foetus undergoes rapid and continuous cellular proliferation that increases the cell population which can be targeted for gene transfer by delivery systems such as integrating lentiviral vectors (David and Peebles, 2008). Proof of somatic gene transfer *in utero* was established in Gunn rats expressing Crigler Najjar disease (Coutelle *et al*, 2005). Successful foetal gene therapy has been demonstrated in both large and small mammalian models including mice and monkeys via intrahepatic, intratracheal and intraamniotic methods (Porada *et al*, 1998; Walsh, 1999; Themis *et al*, 2005; Wagner *et al*, 2009). Tarantal and coworkers demonstrated successful gene transfer to rhesus monkeys *in utero* (Jimenez *et al*, 2005; Tarantal *et al*, 2005). MLV and HIV-1 based vectors were administered to fetal Rhesus monkeys by intraperitoneal and intrahepatic routes. Results showed prolonged transduction and expression of transgenes in multiple organ systems. Seppen *et al*, (2003) investigated the effect of *in utero* injection of lentiviral vectors carrying the hepatic enzyme bilirubin UDP-glucuronyltranferase 1A1 (UGT1A1) which is absent in Crigler Najjar type 1 patients. Results showed sustained expression of trangenes and correction of the UGT1A1 deficiency. Several other studies have also reported the correction of metabolic disorders, coagulapathies and neurological disorders using adenoviral and retroviral vectors. (Mackenzie *et al*, 2002; Rucker *et al*, 2004; Waddington *et al*, 2004).

Stable and efficient transduction of several tissues including pulmonary and heart cells can be achieved when using RV vectors. This supports the potential for fetal gene therapy to reach organs that may be inaccessible after birth. Gregory *et al*, (2004) also demonstrated the ability of non-primate lentivirus (EIAV) to transduce murine skeletal muscles *in vivo*.

The foetal immune system is not fully developed until after birth therefore foetal models will induce lifelong immune tolerance to transgenes expressing transgenic proteins delivered by viral vector constructs. Waddington *et al*, (2004) reported the correction of human factor IX deficiency in hFIX knock out models when using a HIV-1 based lentiviral vector. Vector particles were injected into the fetal liver and plasma FIX antigen levels were increased to 5% above normal human plasma levels. Importantly, lentiviral mediated delivery of hFIX failed to induce an immune response.

Unfortunately the high frequency of cell division in the developing fetus may increase the risk of IM. In addition, caution is required that no vector transmission occurs to the germ line or mother by vector particles (Schneider and Coutelle, 1999, Wagner *et al*, 2009). Lee *et al*, (2007) demonstrated the transduction of a subpopulation of gonadal cells following *in utero* administration of lentiviral vectors highlighting the potential for germline transmission. The semi random insertion pattern of RV's could also have disastrous effects on foetal gene therapy with IM of foetal cells resulting in the formation of malignant tumours.

Key genes involved in processes such as cell cycle metabolism and differentiation are also highly expressed in the foetus therefore the risk of integration into cancer targets is increased. Indeed, the potential for carcinogenesis induced by IM of viral vectors *in utero* highlights the safety issues associated with *in utero* gene transfer; however this feature also makes the foetus a potentially useful genotoxity model. The high levels of gene expression and open chromatin configuration of much of the genome in the foetus can be exploited to predict the genotoxic potential of integrating viral vectors such as retroviruses. In addition the use of foetal gene transfer in genotoxicity studies will improve understanding of foetal development and oncogenesis. Themis *et al*, (2005; 2012 unpublished data) demonstrated a high incidence tumour development in foetal models following lentiviral mediated gene transfer (2005; 2012 unpublished data). These studies show that the foetus model is highly sensitive to vector associated genotoxicity. Moreover, the integration patterns of lentiviruses *in utero* support the use of these models to improve our understanding of viral mediated pathogenesis and IM.

1.1.8 Modifications to the design of RV vectors to avoid insertional mutagenesis

Several models have focused on the modification of RV and LV vectors. Modifications to the design of retroviral vectors have been used to increase the transduction efficiency of T lymphocytes and selective elimination of RV marked cells containing the suicide gene Herpes simplex virus thymidine kinase (*HSVtk*). An immune reaction is often mounted against allogenic T lymphocytes in an event known as Graft versus host disease (GVHD). *HSVtk* modified cells are able to eliminate T lymphocytes in the event of GVHD. Fehse *et al* (2002) recognised the use of a novel 'sort – suicide' fusion gene vector for transfusion of T cells. The inclusion of a truncated human CD34 marker and a splice corrected version of the *HSVtk* transgene improved the positive selection and elimination of gene modified cells respectively. Hence, the use of suicide genes in gene therapy vectors may allow removal of cells that develop clonally dominance. This mode of safety, however, removes all vector infected cells not only unwanted cell populations and thus eradicates the gene therapy treatment.

Several studies have reported the inclusion of chromatin insulators such as chicken β globulin 5' DNase I hypersensitive site 4 (cHS4) into the U3 region of the LTR (Ramezani *et al*, 2008). RV and LV vectors are frequently subjected to position effects produced by host chromosomes (Ramezani *et al*, 2003). This often results in transcriptional silencing of transgenes. Insulator sequences are genetic elements that are located close to chromatin domain boundaries (Emery *et al*, 2000). CHS4 insulators can repress transcriptional activation generated by enhancers or reverse the silencing effects exhibited by heterochromatin regions (Li *et al*, 2009). These sequences can also act bi-directionally to suppress both promoter and enhancer activity and achieve high levels of titre production and transgene expression (Emery *et al*, 2000; Ramezani *et al*, 2003;2008; Nienhuis *et al*, 2006). Ultimately, the inclusion of chromatin insulators can reduce the rate of gammaRV/LV vector-mediated genotoxicity and malignant transformation (Ryu *et al*, 2007; Li *et al*, 2009).

Pseudotyping of viral vector constructs is another effective method of reducing the risk of IM and has been applied to genotoxicity models (Ramezani *et al*, 2008; Mátrai *et al*, 2010). Species-specific interaction between the viral envelope proteins and host cell receptor can restrict viral entry into specific target cells. Pseudotyping involves substitution of the wild type virus envelope glycoproteins for heterologous envelope sequences. This can greatly improve the range of host cells infected by RV/LV particles (tropism) (Kay *et al*, 2001; Cronin *et al*, 2005) and transduction efficiency. Wong *et al*, (2003) developed an EIAV based vector pseudotyped with the Rabies virus envelope glycoprotein. This study demonstrated that pseudotyped EIAV vectors were able to produce high titre vectors and successfully transduce striatum cells within the central nervous system.

Other modifications include the use of ubiquitous and eukaryotic tissue specific promoters that are used to replace viral promoters to drive transgene expression and improve transduction efficiency (Sivalingham and Kon, 2011). These promoters prevent the expression of genes that could be harmful to certain cell types. Hence, expression of cytokines can be restricted to cells of the immune system rather than to the liver and therefore not cause unwanted hepatocyte proliferation.

1.1.9 Hypothesis

Lentiviral vector constructs were administered to foetal mice (MF-1 strain) *in utero* at 16 days gestation. The first objective of this study is to test the hypothesis that cancer development occurred as a result of non-primate LV integration into the host genome.

We previously demonstrated a correlation between the insertion patterns of EIAV and FIV vectors and the development of HCC in foetal models (Themis *et al*, 2005). Therefore in this study it is hypothesised that EIAV and FIV vectors would integrate into RefSeq genes associated with cancer and yield consistent vector copy number in samples infected at the same dose.

This current study also tested the hypothesis that host intrinsic factors such as gene transcription activity, integration site and vector copy number can influence clonal dominance and the development of oncogenesis in foetal mouse models.

It was predicted that integration sites retrieved from these foetal models would be useful for constructing gene expression profiles and oncogeneic networks associated with LV integration.

2.1 MATERIALS

The materials used in these experiments are listed below, along with details of the suppliers from which they were purchased.

2.1.1 General chemicals

Chemical/reagent	Company name
Agarose	Fisher Scientific (Loughborough, UK)
BSA (Bovine Serum Albumin)	Sigma Aldrich (Dorset, UK)
Chloroform	Fisher Scientific
EDTA (ethylene diamine tetraacetic acid)	
Double distilled water (ddH_2O)	Autoclaved Purite water[1]
DMSO (dimethyl sulfoxide)	Sigma-Aldrich
Ethanol	Hayman LTD (Essex, UK)
Ethidium bromide	Fisher Scientific
Glycerol	
IMS (industrial methylated spirit)	Hayman LTD
Isopropanol	Fisher Scientific
Phenol	Sigma-Aldrich
Potassium chloride	Fisher Scientific
SOC medium	Qiagen (West Sussex, UK)
Sodium chloride	Fisher
Sodium citrate	Sigma Aldrich
Sodium hydroxide	BDH
Tris Borate Acid (TBE)	Fisher Scientific

Table 4. General chemicals and reagents used in experiments [1] represents ddH_2O.

2.1.2 Transformation reagents

Chemical/reagent	Company name
ddH_2O	Qiagen
pDrive cloning vector (50 ng/µl)	
2X Ligation master mix	
QIAGEN EZ Competent Cells	
SOC medium	
Tryptone	Fisher Scientific
Yeast extract	Fisher Scientific
Sodium Chloride	
Agar	
Ampicillin	
X-gal (5-bromo-4-chloro-3-indolyl β-D-galactoside)	
IPTG (Isopropyl β-D-1-thiogalactopyranoside)	

Table 5. Summary of reagents required for transformation protocols.

2.1.3 Ligation amplification mediated (LAM PCR) reagents

Chemical/reagent	Company name
Linear PCR/First and second round PCR	
ddH_2O	[1]
10X PCR buffer containing magnesium	Kappabiosystems (Woburn, MA, USA)
dNTP mix (deoxyribonucleotide triphosphates)	Invitrogen (Paisley, UK)
Kapa *Taq* DNA Polymerase (Kappabiosystems)	KapaBiosystems
DNA capture of magnetic beads	
Dynal beads	Invitrogen
Washing solution containing 1X PBS^ & 0.1% BSA	Sigma Aldrich (PBS) & Fisher (BSA)
Binding solution	Invitrogen
Second strand synthesis	
Klenow (Large Fragment DNA Polymerase I	Invitrogen
Klenow dilution buffer	
dNTP's	
10x React2Buffer	
10x Hexanucleotide mix	Roche (West Sussex, UK)
Restriction digest of the double stranded DNA with Tsp509I	
ddH_2O	[1]
10x NEB buffer	New England Biolabs (Herts, UK)
Tsp509I	
Ligation of linker cassette	
Ligation casette	Sigma Aldrich
10mM ATP	Epicentre Biotechnologies (Post Rd, Madison, USA)
10X Fast Link ligation buffer	
2U/µL Ligase	
Agarose gel electrophoresis of LAM PCR products	
3% agarose gel	
6X Agarose gel loading dye	Fisher Scientific
TrackIt™1Kb DNA ladder	Invitrogen
1Kb Plus DNA ladder	

Table 6. Summary of reagents used in LAM PCR analysis. [1] is used to denote ddH_2O that was prepared from autoclaved water.

2.1.4 Oligonucleotides used for LAM PCR analysis

Gene name	Annealing Temperature	Sequence
Biotinylated linear PCR		
EIAV	60°C	GGCCAGGAACACCTCCAG
EIAV	60°C	GCTGTCCCGGGGATCCTACG
FIV-LTR-5A	60°C	GTTCTCGGCCCGGATTCC
FIV-LTR-5B	60°C	CCCGGATTCCGAGACCTC
3LTR Linear HIV A	60°C	GAGCTCTCTGGCTAACTAGG
3LTR Linear HIV B	60°C	GAACCCACTGCTTAAGCCTCA
Linker sequences		
LC1		GACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG
LC3		AATTCCTAACTGCTGTGCCACTGAATTCAGATC
First round PCR		
EIAV-LTR-5C	60°C	GCGCCCCTCTCAGGTCCC
EIAV LC-1	60°C	GACCCGGGAGATCTGAATTC
FIV-LTR-5C	60°C	CTCGACAGGGTTCAATCTC
FIV LC-1	60°C	GACCCGGGAGATCTGAATTC
HIV-LTR-5C	60°C	AGCTTGCCTTGAGTGCTTCA
HIV LC-1	60°C	GACCCGGGAGATCTGAATTC
Second round PCR		
EIAV-LTR-5D	60°C	CCCTGTTCGGGCGCCAAC
EIAV-LCIII	60°C	AGTGGCACAGCAGTTAGG
FIV-5LTR-5D	60°C	CTCAAAAGTCCTCAACAAAG
FIV LCIII	60°C	AGTGGCACAGCAGTTAGG
HIV-3LTR-5D	60°C	AGTAGTGTGTGCCCGTCTGT
HIV LCIII	60°C	AGTGGCACAGCAGTTAGG

Table 7. List of primers used for LAM PCR analysis.

2.1.5 RNA extraction for RT PCR and QPCR analysis

Chemical/reagent	Company name
TRI® reagent	Sigma-Aldrich
Chloroform	Fisher
2-propan-1-ol (isopropanol)	Sigma-Aldrich
75% ice cold ethanol	See 1.1.22

Table 8. Reagents used for isolation of total RNA from tissue samples.

2.1.6 Reagents for *DNase I* treatment

Chemical/reagent	Company name
10X reaction buffer	Sigma-Aldrich
Amplification Grade DNase I (1,000 units)	
Stop Solution	
70% ethanol	See 2.2.21

Table 9. General reagents used for *DNase I* treatment.

2.1.7 Reagents: CDNA synthesis for RT (Reverse Transcriptase) - PCR analysis

Chemical/reagent	Company name
Oligo (dT) 20 primer	Invitrogen
dNTP mix	
DEPC-treated water	
10X RT buffer	
$MgCl_2$	
0.1 M DTT	
RNaseOUT. (40 U/µl)	
SuperScript. III RT (200 U/µl)	
RNase H	

Table 10. General reagents used for CDNA synthesis of samples used for RT PCR.

2.1.8 RT PCR reagents

Chemical/reagent	Company name
10x PCR buffer	Invitrogen
50mM $Mgcl_2$	
10mM dNTP's mix	
2.5U Platinium *Taq* DNA polymerase	

Table 11. General reagents used for RT PCR.

2.1.9 Oligonucleotides for RT-PCR analysis

Gene name	Annealing temperature	Sequence
GAPDH	59ºC	Forward - GGCCTTGACTGTGCCGTTGAATTT Reverse - ACAGCCGCATCTTCTTGTGCAGTG
Coro7	59ºC	Forward - AGGCATGTGACCTACCTTGG Reverse - CCGAAGCTGCTTGTCCTTAC
Nek9	63ºC	Forward - CAGGTGGAGGTTGACAGGTT Reverse - CTCTTCAAGGTGGCTCAG
Cx39	63ºC	Forward - TGCTAGTGAGGCCACTTCCT Reverse - CATGTCATTCCACCCCTACC
Sac	63ºC	Forward - CCTCCTGGTGAAAAAGCAG Reverse - TTCTCTGTCTCCAGGCGTT
Pcd1lg2	63ºC	Forward - TGTGCTGCCTTTTCTGTGTC Reverse - GCCAGGACACTTCTGCTAGG
Tnfrsf19	63ºC	Forward - CGCTGCCATTCTCTTCCTAC Reverse - GGTCAAAGCATGACAGCTCA
Similar to cyclin fold protein 1	63ºC	Forward - TATCTGCACCTGAGCACCAG Reverse - CCTCTTTGCCTCTTCCACAG

Table 12. Oligonucleotides used RT PCR amplifications.

2.1.10 RNA purification for Real -Time Quantitative (Q) – PCR reagents

Chemical/reagent	Company name
10X reaction buffer	Sigma-Aldrich
Amplification Grade DNase I (1,000 units)	
Stop Solution	
70% ethanol	See 1.1.22
β-ME (2-Mercaptoethanol)	Agilent technologies, Stratagene (Edinburgh, UK)
Lysis solution	
1x low salt wash solution	
Elution buffer	

Table 13. General reagents/chemicals used for QPCR reactions.

2.1.11 CDNA synthesis for Real -Time Quantitative (Q) - PCR reagents

Chemical/reagent	Company name
ddH_2O	[1]
10X RT Buffer	Applied Biosystems
10X RT Random Primers	
25X dNTP mix (100mM)	
MultiScribe™, Reverse Transcriptase, 50 U/µL	
RNase Inhibitor	

Table 14. General reagents/chemicals used for CDNA synthesis of total RNA for QPCR reactions

2.1.12 Q-PCR reagents for gene expression analysis and vector copy number (VCN)

Chemical/reagent	Company name
ddH_2O	1
2x TaqMan Universal PCR Master Mix containing AmpliTaq Gold®DNA polymerase, Amperase UNG, dNTP's and dUTP	Applied Biosystems
20x TaqMan gene expression assay	

Table 15. General reagents/chemicals used for TaqMan PCR reactions.

2.1.13 Q-PCR primer probe sets for VCN analysis

Chemical/reagent	Company name
***GAPDH* - mouse** (assay ID: Mm99999915_g1) ***GAPDH* - human** (assay ID: Hs0275858991_91) **EIAV packaging signal (psi)** *Forward sequence:* ATTGGGAGACCCTTTGACATTG *Probe:* AGCAAGGCGCTCAAGAAGTTAGAGAAGGTG *Reverse sequence:* ACCAGTAGTTAATTTCTGAGACCCTTGTA **WPRE** *Forward sequence:* TTCTCCTCCTTGTATAAATCCTGGTT *Probe:* CTGTCTCTTTATGAGGAGTTGTGG *Reverse sequence:* CCACGCCACGTTGCCTGACAACGG	Applied Biosystems

Table 16. TaqMan primer probe sets used for VCN studies.

2.1.14 Pre designed and custom made TaqMan probes for gene expression analysis – Applied Biosystems

Gene name	Assay ID	Exons amplified	Amplicon product size
18sRNA	4310893E	N/A	187
GAPDH	Mm99999915_g1	2 – 3	107
Pah	Mm01224032_m1	2 – 3	68
Acvr2a	Mm01331093_m1	1 – 2	64
Setd5	Mm01184707_m1	1 – 2	118
Mrp1	Mm01208141_m1	1 – 2	74
Park7	Mm00498538_m1	3 – 4	92
Uvrag	Mm00724367_m1	3 – 4	132
Mrpl23	Mm01158967_m1	1 – 2	127
Rabgef1	Mm00727426_s1	8-9	66
Coro7	Mm00504152_m1	9 – 10	103
Nek9	Mm01307401_m1	21 – 22	70
Pscd3/Cyth3	Mm00441015_m1	1 – 2	90
Tnfrsf19	Mm00443506_m1	4 – 5	96
CX39/Gjd4	Mm00462088_m1	1 – 2	72
Pcd1lg2	Mm00451734_m1	4 – 5	95
Foxa2	Mm00839704_Mh	2 – 3	64
Hnf1A	Mm00493434_m1	1 – 2	134
Hnf4A	Mm00433964_m1	8 – 9	114

Table 17. TaqMan probes and unlabelled oligonucleotide sequences used for TaqMan analysis of lentivirus insertion genes.

2.1.15 Tissue culture reagents

Chemical/reagent	Company name
DMEM (Dulbecco's modified Eagle's medium) containing GlutaMax™, 1000 mg/L and sodium pyruvate	Fisher Scientific
Fetal bovine serum	
Pen/strep (penicillin/streptomycin)	
10X Trypsin-EDTA (containing 0.5% trypsin in 5.3mM EDTA)	
DMSO (dimethyl sulfoxide)	Sigma-Aldrich
1X PBS^	
Trypan blue	GIBCO™ Invitrogen
Virkon disinfectant	Fisher Scientific

Table 18. General reagents used for tissue culture.

2.1.16 Reagents for DNA extraction from frozen tissue samples and cultured cells

Chemical/reagent	Company name
Phenol	Sigma-Aldrich
Chloroform	Fisher Scientific
Extraction buffer	See 1.1.22
Proteinase K	Fisher Scientific
RNase A	Fisher Scientific
Ethanol	Hayman LTD
70% ethanol	See 2.2.21
Liquid nitrogen	

Table 19. Reagents used for DNA extraction of frozen tissue and cultured HepG2 cells lines.

2.1.17 Reagents for Global methylation assay using Imprint ® Methylated DNA quantification kit

Chemical/reagent	Company name
10x Wash Buffer	Sigma-Aldrich
DNA Binding Solution	
Methylated Control DNA (50 ng/µL)	
Block Solution	
Capture Antibody	
Detection Antibody	
Developing Solution	
Stop Solution	

Table 20. General reagents used for global methylations assays.

2.1.18 Reagents for Quantification of DNA methyltransferase activity using QPCR

Gene name	Assay ID	Exons amplified	Amplificon product size
18sRNA	4310893E	*	*
Dnmt1	Mm00599763_m1	1-2	68
Dnmt3a	Mm00432870_m1	6-7	75
Dnmt3b	Mm00599800_m1	1-2	61

Table 21. TaqMan gene expression assays used to quantify DNA methyltransferase activity

2.1.19 Human cancer cell line

HepG2 (Human hepatocellular carcinoma cell line) was used for mammalian tissue culture. These cells were kindly provided by Dr Amanda Harvey (Brunel University, Uxbridge, UK).

HepG2 cells have been isolated from human liver cancer patients. These cells are frequently used in Genotoxicity studies and identification of reactive components (Knasmuller *et al*, 2004).

2.1.20 EIAV, FIV and HIV vectors

The SMART2Z, SMART2ZWΔ, SMART2hFIX, pLIONhAATGFP and HR'SIN-cPPT-S-FIX- vectors used in this study were originally provided by Dr Themis. These vectors were used to inject fetal mice (via the fetal yolk sac) at 16 days of gestation. Observations showed the development of lentiviral associated oncogenesis following transduction with EIAV vectors. The effect of FIV infection was also explored in this study using FIV based vectors produced using methods described by Condiotti *et al* (2004).

2.1.21 Compositions of buffers and solutions

General buffers & solutions

DNA extraction buffer

50 mM Tris pH 8.0

100 mM EDTA pH 8.0

100 mM NaCl

0.1% SDS

5X TBE (tris-borate EDTA) buffer

500ml ddH_2O

27g Tris base

137.5g boric acid

10ml 0.5M EDTA pH 8.0

1X TBE

700ml ddH_2O

300ml of 5XTBE solution

75% Ethanol

750ml ddH_2O

250ml of Absolute ethanol solution

70% Ethanol

700ml ddH_2O

300ml of Absolute ethanol solution

2.2 METHODS

2.2.1 Quantification of nucleic acids

Nucleic acids (dsDNA, cDNA and RNA) were quantified and absorbance values measured at several wavelengths (260, 280 and 260:280) using the Nanodrop spectrophotometer. The 260 absorbance reading was used to determine the concentration of nucleic acids present in uninfected, normal and tumour mouse samples. The 280 absorbance reading was used to detect protein contamination in mouse samples. ddH_2O/TE nucleic acids buffer was used as a reference sample. 1µl of this reference sample was applied to the Nanadrop using a sterile pipette and this was used to read a zero absorbance for the 'blank'. 1µl of DNA, plasmid DNA, cDNA and RNA samples were then measured using the Nanadrop. For pure DNA and RNA samples the 260/280 ratio given were approximately 1.8 and 2.0 respectively.

2.2.2 Agarose gel electrophoresis

Agarose gel electrophoresis allows negatively charged nucleic acids (DNA and RNA) to be separated according to size. An electrical current is passed through the agarose gel allowing the movement of small molecules. Agarose gel electrophoresis was used to determine the purity and integrity of nucleic acids. A 60ml agarose gel was prepared by dissolving agarose powder in 60ml of 1X TBE solution. The percentage of agarose powder used varied according to the expected size of products being separated. The solution was heated in a microwave for 2 minutes to dissolve the agarose and yield a clear viscous solution. The solution was gently swirled to mix and allowed to cool to approximately 50ºC. 6µl ethidium bromide (10mg/ml stock) was added to the 60ml heated agarose solution before pouring into a casting tray containing a well-forming comb. The ethidium bromide was used to allow flourescent visualisation of nucleic acids under UV light. Once set the comb was removed and the gel immersed (~5mm above the gel) in 1X TBE solution. Loading buffer (Fisher Scientific Cat # BPE633-5) was added and mixed with samples on a parafilm sheet or sterile tubes before loaded into wells on the gel. A separate well was used to load a standard DNA marker (Invitrogen Cat # 10787-018 & 10488-072)

or (Fisher Scientific Cat # F-303SD). This well was used to measure the size of DNA, RNA and cDNA fragments ranging from +/- 100 to 12,000 base pairs. A power pack was attached to the gel tank using a positive (anode) and negative (cathode) electrode. The power was switched on to allow separation of nucleic acids. A typical gel was allowed to run at 50V for aproximately 1 hour. However, this time varied according to the experiments required to allow clear seperation of nucleic acids. To view results the power pack was switched off and the 1XTBE solution decanted into a destaining bag. DNA, RNA and cDNA fragments were visualised using ultraviolet illumination of the gel. Gel images were captured and printed on to thermal paper using the Alpha Imager 2200 device (Alpha Innotech Corporation).

2.2.3 Preparation of LB agar plates and LB medium

Bacterial cell including transformed competent cells were grown on plates enriched in LB agar. LB agar was prepared in a sterile bottle using the reagents and quantities listed in Table 22.

Chemical/reagent	Quantities/volume	LB agar	LB medium
Tryptone	5g	✓	✓
Yeast Extract	2.5g	✓	✓
Sodium Chloride	2.5g	✓	✓
Agar	7.5g	✓	
Distilled water	100ml	✓	✓

Table 22. Reagents /chemicals required for preparation of LB agar plates and LB medium.

The LB mixture was autoclaved and the solution cooled down in a 50ºC water bath. All reagents were transferred to a sterile air flow hood. 250µl of ampicillin (100µg/L) was added to 500ml of LB agar solution. The solution was mixed by swirling and the solution poured into sterile plates making sure to avoid air bubbles. LB agar was left to solidify then stored in a 37ºC incubator until ready to use. 40µl of X-gal (40mg/ml) and 40µl (100mM) of IPTG was aliquoted and spread on to each LB agar plate using a sterile glass spreader immersed in IMS. LB agar plates were then left to dry at 37ºC until ready for use.

2.2.4 Transformation of PCR products

PCR products were cloned into the pDrive cloning vector (Qiagen, Cat # 231122; Figure 4). A ligation master mix was prepared and placed on ice. The volumes required for one reaction consisted of 1µl of pDrive Cloning Vector, 2µl of LAM PCR product, 2µl of distilled water and 5µl of ligation master mix. Each reaction mixture was mixed by reverse pipetting and incubated in ice for 2 hours. Tubes containing Qiagen EZ competent cells (Qiagen Cat # 231222) were briefly thawed on ice. 2µl of ligation-PCR master mix was then added to competent cells and mixed by gently flicking the tube. The transformation mixture was incubated in ice for 5 minutes then placed in a 42ºC water bath for 30 seconds. Transformation mixtures were then transferred to ice and incubated for 2 minutes. After incubation 250µl SOC medium was thawed at room temperature then added to each transformation mixture. 100µl of reaction mixtures were transferred to freshly prepared LB agar plates (Table 22) using a pipette and evenly spread using a sterile glass spreader. Samples were left on the plates for 10min before being stored in a 37ºC incubator overnight.

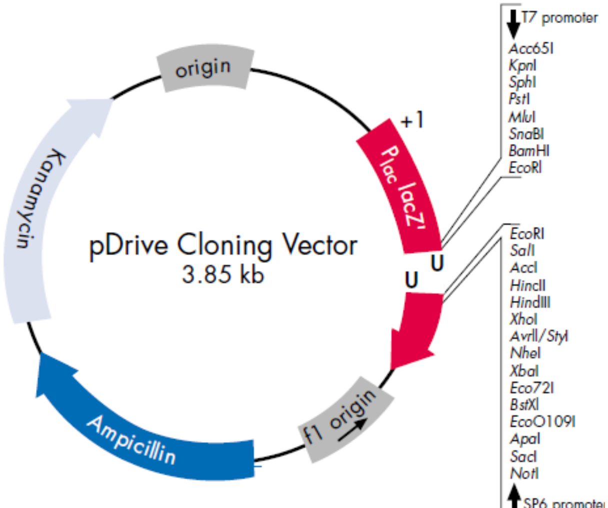


Figure 11. Schematic representation of the pDRIVE vector used to clone PCR products and transform bacterial cells prior to DNA sequencing. The cloned plasmids were purified. 200-300ng of sample was sent to the Dundee DNA Sequencing service (MRCPPU, College of Life Sciences, University of Dundee, Scotland, www.dnaseq.co.uk) to determine the nucleotide sequences of PCR products. Image taken from Qiagen® PCR Cloning Handbook (2001).

2.2.5 Selection of transformed clones using 'blue/white screening'

The LB plates were stored in the fridge for approximately 3 hours (4ºC) to allow the formation of blue/white screening of colonies. Blue/white screening is used to differentiate between transformed and untransformed bacteria cells and identify desired clones. Untransformed bacteria are unable to grow on LB agar plates containing ampicilin as they lack the antibiotic resistance gene present in the pDRIVE vector. Recombinant plasmids confer antibiotic resistance and are able to grow on ampicilin plates. Transformed clones were also selected through the use of blue white screening. The pDRIVE vector carries the LacZ gene which catalyses the breakdown of X-gal. The addition of IPTG to LB plates helped promote LacZ expression. The ligation of a PCR product into the LacZ coding region disrupts LacZ activity and turns colonies white or pale blue on LB plates. Untransformed clones that lack inserted sequences were observed as dark blue colonies on LB plates.

2.2.6 Growth of transformed plasmids in LB medium

Fresh LB medium was prepared using the reagents and quantities listed in Table 24. 100µl of ampicillin (50µg/ml) was added to 100ml of LB medium under sterile conditions. The LB medium was then mixed by swirling. 5ml of LB medium was added to 50ml Falcon tubes. Single transformed colonies were picked from the plate using sterile picks and transferred to respective 50ml Falcon tubes. Tubes were then incubated in a 37ºC shaker overnight.

2.2.7 Storage and recovery of bacterial cells for sequencing analysis

250µl of incubated bacterial culture was added to 250µl of 20% glycerol solution and the mixture vortexed to mix. Reaction tubes were stored in a -80ºC freezer overnight. Frozen glycerol stocks were extracted using a sterile rod and transferred to 50ml tubes containing 5ml of fresh LB medium and ampicilin (Table 22). The LB medium was incubated in a 37ºC shaker overnight. This bacterial culture was used to prepare plasmid DNA for sequence analysis.

2.2.8 Plasmid DNA purification

The QIAprep Spin Miniprep Kit (Qiagen Cat # 27106) was used to purify plasmid DNA.

LB broth containing bacterial cultures were spun down in a centrifuge for 5 minutes. Pellets (competent cells) were re-suspended in 250µl of re-suspension buffer (P1) containing RNaseA. The reaction mixture was mixed thoroughly by reverse pipetting followed by vortexing then transferred to a 1.5ml centrifuge tube. 250µl of Lysis buffer (P2) was added to each reaction and mixed by inverting tubes 6 times until a clear viscous solution was formed. 350µl of Neutralising buffer (P3) was added to each tube and the reaction mixture quickly mixed by inverting the tube 10 times. The cloudy solution was centrifuged at 13,000rpm for 10 minutes to yield a white pellet. The supernatant was transferred to QIA prep spin column using a pipette. The reaction mixture was re-centrifuged for 1 minute and the flow through solution removed using a pipette. 0.5ml of Binding buffer (PB) was added to each spin column and the tubes were re-centrifuged for 1 minute. The flow through was removed and 0.75ml of wash buffer (PE) added to each column. Centrifugation (1 minute) and removal of flow through solution was repeated twice to remove traces of ethanol derived from PE buffer. The QIA prep column was then transferred to a sterile 1.5ml centrifuge tube. 50µl of Elution buffer (EB) containing 10Mm of Tris-CL (pH 8.5) was added to the centre of each spin column and left to incubate at room temperature for 1 minute. The purified plasmid DNA was eluted from reaction tubes by centrifugation at 13,000rpm for 1 minute using a bench top centrifuge.

2.2.9 Restriction digestion of purified plasmid DNA

The pDrive vector was digested using the restriction enzyme *EcoRI*. *EcoRI* restriction sites were located between 306-317 bp and 318-322bp of the pDrive vector. A restriction digestion master mix was prepared using the reagents and quantities listed in Table 23. Control reaction mixtures consisted of undigested plasmid DNA.

Reagents	Volumes x1 (µl)
ddH_2O	17.5
10x NEB buffer 4	2
EcoRI	0.5
DNA (purified plasmid)	2
Total	20

Table 23. Reagents used for restriction digestion of plasmid DNA using *EcoRI.*

The reaction mixture was mixed by reverse pipetting followed by brief centrifugation at 13,000rpm. Restriction digestion reactions were incubated for 1 hour at 65°C to ensure complete digestion of plasmid DNA. Restriction digested and control reaction samples were loaded and run on a 1% agarose gel.

2.2.10 Colony PCR

Colony PCR was used to screen for plasmid inserts. A PCR was performed using primers specific to the SP6 promoter and M13 regions present in the pDrive cloning vector (Figure 11). A PCR master mix was prepared using the reagents and quantities listed in Table 24.

Chemical/reagent	Working volumes x1 (µl)
ddH_2O	7.6
10X PCR buffer containing Mg2+ (KappaBiosystems)	1
10mM dNTP	0.2
SP6 forward primer	0.1
M13 reverse primer	0.1
Taq polymerase (1:20 dilution)	1

Table 24. Reagents and volumes required to prepare a one times reaction mixture for colony PCR.

A 1 in 100 dilution was made of the LB medium containing transformed cells (LB broth). 10µl of PCR master mix was added to a 0.5µl of diluted LB broth mixture in a 0.5ml centrifuge tube. A few drops of mineral oil were added to the top of the reaction mixtures to prevent sample evaporation.

The PCR reaction tubes were then transferred to a thermal cycler (Hybaid Ominigene thermal cycler) and the colonies amplified using the parameters listed in Table 25.

Step	Temperature	Time	No. cycles
Denaturation	95°C	5 minutes	1
Denaturation	95°C	1 minutes	40 x
Annealing	54°C	1.5 minute	
Extension	72°C	90 seconds	
Extension	72°C	5 minutes	1

Table 25. PCR parameters used for colony PCR.

The amplified PCR products were loaded onto a 1% agarose gel and ran at 60V for 45 minutes. The presence of bands confirmed the presence of transformed competent cells containing PCR products.

2.2.11 DNA sequencing

The Nanodrop was used to measure the absorbance values of purified plasmids. Mean absorbance values were calculated and samples diluted with water to 300ng in a total volume of 30µl. The labeled samples were delivered to Dundee Sequencing Service for sequence analysis. DNA sequencing was carried out by DNA Sequencing & Services (MRCPPU, College of Life Sciences, University of Dundee, Scotland, www.dnaseq.co.uk) using Applied Biosystems Big Dye Ver 3.1 Chemistry on an automated capillary DNA sequencer (Applied Biosystems DNA analyser Model # 3730).

2.2.12 LAM PCR

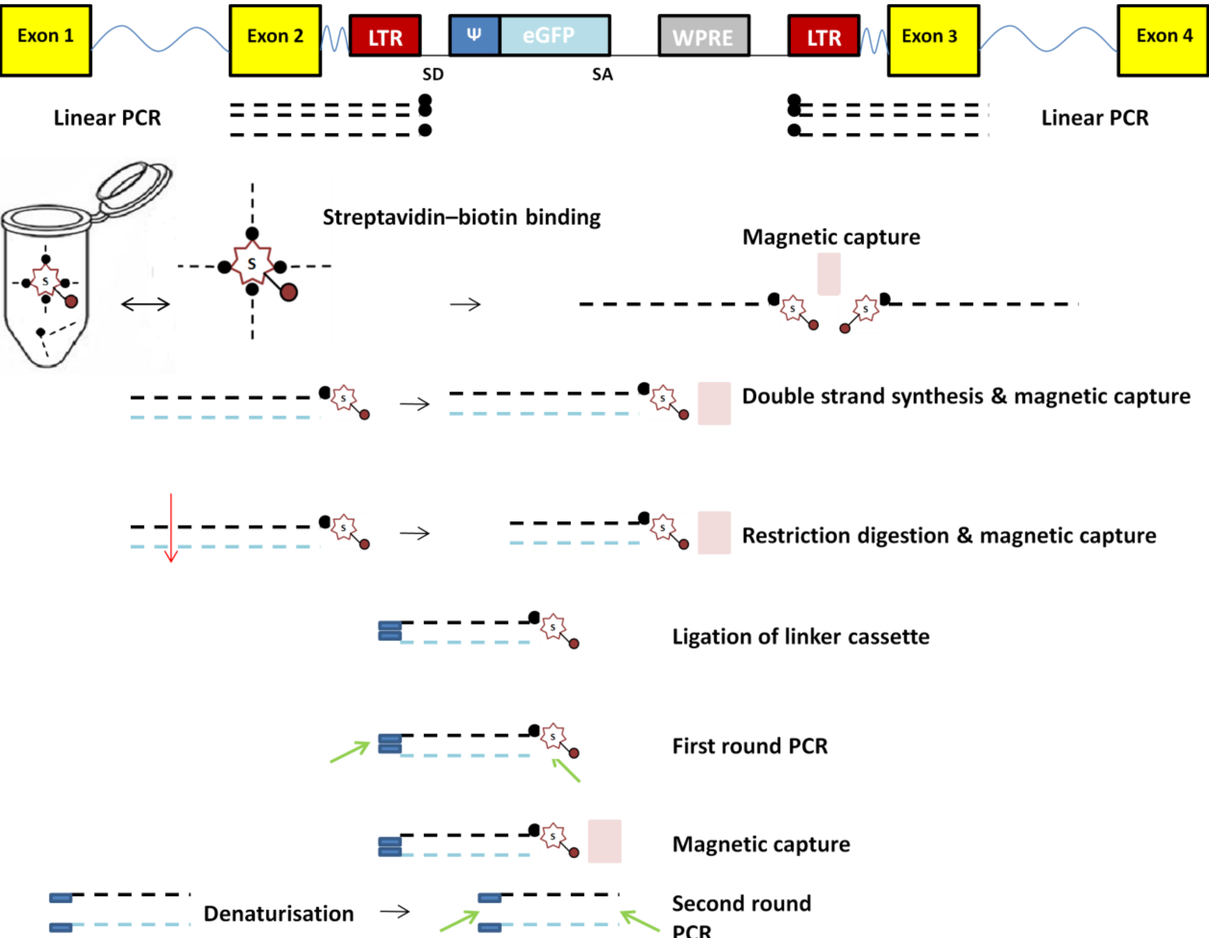


Figure 12. A schematic overview of the steps required for LAM PCR. The proviral-genomic DNA junction is first amplified using biotinylated – LTR specific primers. This generates a linear PCR product complementary to the provirus insertion site starting from the LTR region into the neigbouring host DNA. Streptavidin Dynabeads are added to a sterile tube containing linear (biotinylated) PCR products. This results in streptoavidin-biotin binding. Streptoavidin-biotin complexes are isolated using magnetic capture. Klenow fragment (*DNA polymerase 1*) is used to synthesise a strand complementary to linear PCR products containing proviral and flanking mouse genomic sequences. This is followed by magnetic capture. *Tsp509I* (a 4 cutter enzyme) was used to digest double stranded provirus junction at 5′- AATT – 3′ restriction sites followed by magnetic capture of the resulting products. A known linker cassette sequence was ligated onto the end of isolated products. PCR and nested PCR amplification was then used to amplify the linker-genomic DNA-provirus complex using virus and linker specific primers.

Linear PCR

Linear amplification reactions were conducted on FIV infected and uninfected tissue samples using biotinylated primers that hybridise to the LTR sequence of the FIV lentiviral constructs (Table 7; Figure 12). A PCR master mix (Table 26) was prepared in a 1.5ml centrifuge tube and placed on ice. The PCR master mix was briefly centrifuged at 13,000 rpm (17,900 x g) using a table-top centrifuge and aliquoted into 0.2ml centrifuge tubes for each reaction. 2µl of DNA was then added to each respective tube.

Chemical/reagent	**Final concentration**	**Working volume x1(µl)**
Distilled water	-	42.5
PCR buffer + magnesium mix	x1	5
dNTP solution	0.2mM	1
FIV-LTR-5A primer	2.5nM	0.5
FIV-LTR-5B primer	2.5nM	0.5
KAPA Taq DNA Polymerase	2.5U	0.5
DNA template (liver tissue)	50ng/µl	2
Total Volume		50

Table 26. Reagents and volumes required to make a linear PCR master mix for one reaction.

The PCR sample tubes were transferred to a thermal cycler (MJ Research PT-225 Tetrad Therma cycler) and amplified for 3 hours and 58 seconds. The PCR parameters are listed in Table 27.

Step	**Temperature**	**Time**	**No. cycles**
Denaturation	95°C	5min	1x
Denaturation	95°C	1 min	50 x
Annealing	60°C	45 sec	
Extension	72°C	90 sec	
Extension	72°C	10 min	1x
Pause	4°C	-	-

Table 27. PCR parameters required for linear PCR of lentiviral infected samples.

Once the linear PCR amplifications were completed 0.5µl of Platinum *Taq polymerase* was added to each PCR reaction tube. PCR tubes were then re-amplified

using the same parameters listed in Table 27 in order to increase the number of biotinylated products containing provirus insertions sites. PCR products were then stored at 4ºC overnight.

Preparation of the linker cassette

The linker cassette contains short sequences (5'-AATT-3') with a *Tsp509I* restriction enzyme site. This synthetic linker cassette sequence was prepared using LC1 and LC3 oligo sequences (Table 7). A linker cassette master mix was prepared using 40µl of LC1 (final concentration of 20µM), 40µl of LC3 (final concentration of 20µM), 55µl of Tris pH 7.5 (137.5mM) and 20µl of $MgCl_2$ (5mM) and heated in a hot block (95˚C) for 5 minutes. The master mix was then left to cool slowly for several hours.

DNA capture of magnetic beads

Biotin binds to streptoavidin with an extremely high affinity. This was exploited using the Dynal® kilobaseBINDER™ Kit (Invitrogen Cat # 601-01). Streptoavidin Dynabeads were vortexed to re-suspend the beads. 20µl of the beads were added to sterile centrifuge tubes and placed in a magnetic rack to immobilise beads. The supernatant was removed with a pipette and the brown pellet of beads mixed with 20µl of fresh washing solution (0.1% BSA dissolved in 1X PBS^) by flicking the tube. Tubes were then transferred to a magnetic rack. This step was repeated once. Beads were then gently re-suspended by flicking in 20µl of binding solution making sure to avoid foaming. The beads were again immobilised in a magnetic rack, supernatant removed and re-suspended in 50µl of binding solution. Beads were then immobilised and the supernatant removed before conducting a final re-suspension step. The magnetic beads were then re-suspended in 50µl of linear PCR products. Reaction mixtures were incubated for one hour at room temperature to allow the formation of streptoavidin-biotin complexes. Reaction mixtures were kept re-suspended by gently flicking the tubes every 15 minutes.

Second strand synthesis

Second strand synthesis was prepared using the DNA synthesising enzyme klenow (*DNA polymerase I* fragment) and a hexanucleotide mix. A second strand synthesis master mix was prepared in a sterile 1.5ml centrifuge tube on ice and briefly mixed at 13,000rpm. The volumes of the reagents used are shown in Table 28.

Chemical/reagent	**Final concentration**	**Working volume x1 (µl)**
Distilled water	-	14.5
diluted Klenow	2U/µl	1
dNTP	2.5µM	0.5
React2Buffer	1x	2
Hexanucleotide mix	1x	2
Total		20

Table 28. Reagents and volumes required to make for one reaction for second strand synthesis.

Klenow (Large Fragment DNA Polymerase I) was used to dilute to Klenow dilution buffer and placed solution on ice. Incubated linear PCR products were washed with 100µl of double distilled water and immobilised in a magnetic rack. The supernatant was removed and beads re-suspended with 20µl of second strand synthesis master mix. All reactions were incubated in a 37°C water bath for one hour. Samples were kept re-suspended by gently flicking tubes every 15 minutes.

Restriction Digest of the double stranded DNA with Tsp509I

TSP509I was used to digest the double stranded DNA prior to the ligation of the linker cassette sequence. A restriction digestion master mix was prepared using the reagents and quantities in Table 29 and placed on ice.

Stock solutions	**Final concentration**	**Working volume x1 (µl)**
Distilled water	-	17
10x NEB buffer	1x	2
4000µg/ml Tsp509I	200µg/ml	1
Total		20

Table 29. Reagents and volumes required to make for one reaction for restriction digestion using *Tsp509I*.

Second strand synthesis samples were removed from the water bath and placed in a magnetic rack. The supernatant was removed from each sample. Washing, immobilisation and supernatant removal steps were performed firstly with 80µl and repeated with 100µl of ddH_2O. 20µl of reaction digest master mix was added to the concentrated pellet in each reaction tube. All samples were incubated in a 65˚C water bath for one hour. Samples were kept re-suspended by gently flicking every 15 minutes.

Concentration of the linker cassette with a Micron-YM30 concentrator.

The linker cassette master mix was concentrated from 200 to 80µl. The 200µl linker cassette solution was added to the column membrane of the Micron-YM30 concentrator (Millpore, Cat # 42409) and the tube spun in a desktop centrifuge at 13,000rpm for 8 minutes. 20µl of ddH_2O was added to the column membrane then spun at 13,000 rpm for 1 minute in a desktop centrifuge. The column was then turned upside down and placed in a sterile 1.5ml eppendorf tube. Centrifugation at 13,000rpm was repeated for approximately 2 minutes. 60µl of ddH_2O was then added to the concentrated linker cassette volume and the sample tube placed on ice.

Ligation of the linker cassette

A ligation master mix was prepared using the reagents and quantities listed in Table 30.

Chemical/reagent	**Final concentration**	**Working volume x1(µl)**
dH_2O	-	4
Ligation casette	-	2
ATP	1mM	1
Ligation buffer	1x	1
Ligase	0.4U/µl	2
Total		10

Table 30. Reagents and volumes required to make for one reaction for ligation the linker cassette using the ligase enzyme.

The restriction digest samples were removed from the water bath and placed on ice. 80µl of ddH_2O was added to each tube and mixed by flicking. The samples were placed in magnetic rack and the pellets re-suspended in 100µl of ddH_20. Sample tubes were immobilised in a magnetic rack and the supernatant removed.

The concentrated samples were then re-suspend with 10µl of ligation master mix and incubated at room temperature for 30 minutes. Restriction reactions were kept re-suspended by gently flicking every 10 minutes.

First round PCR

A PCR master mix was prepared using the reagents and quantities in Table 31 then placed on ice.

Chemical/reagent	Final concentration	Working volume x1 (µl)
Distilled water	-	31
PCR buffer + magnesium mix	1x	5
dNTP solution	0.2mM	1
FIV-LTR-5C primer	0.5µM	5
LC-1 primer	0.5µM	5
Taq polymerase	0.1U/µl	1
DNA template (ligation product)	-	(2)
Total		50

Table 31. First round PCR: Reagents and volumes required to make a PCR master mix for one reaction.

Primers used were specific to both the 5LTR (FIV-LTR-5C) and the LC-1 linker sequences (Table 7). Ligated linker samples were re-suspended and washed with 90µl of ddH$_2$O then placed in a magnetic rack. The supernatant was discarded with a pipette. DNA was denatured by adding 5µl of NaOH (0.1M) to each pellet. Reaction mixtures were then incubated at room temperature for 10 minutes. 48µl of master mix and 2µl of denatured DNA was added to respective PCR centrifuge tubes and placed on ice. The reaction mixture was briefly mixed in a centrifuge then placed on ice. Tubes were transferred to a pre-set thermal cycler using the PCR parameters in Table 32.

Step	Temperature	Time	No. cycles
Denaturation	95°C	5min	1x
Denaturation	95°C	1 min	30 x
Annealing	60°C	45 sec	
Extension	72°C	90 sec	
Extension	72°C	10 min	1x
Pause	4°C	-	-

Table 32. PCR parameters required for first round PCR.

Second Round PCR

Nested primer sequences were used to amplify the products generated by first round PCR. This step used vector-specific (FIV-5LTR-5D) 5'CTCAAAAGTCCTCAACAAAG 3' and linker specific (LC3) 5'AGTGGCACAGCAGTTAGG 3' primers respectively. A master mix for the second round PCR was prepared using the reagents and quantities in Table 33 then placed on ice.

Chemical/reagents	Final concentration	Working volume x1 (µl)
Distilled water	-	32
PCR buffer + magnesium mix	1x	5
dNTP solution	0.2mM	1
FIV-LTR-5D primer	0.5µM	5
LC-3 primer	0.5µM	5
KAPA Taq polymerase	0.1U/µl	1
DNA template (First round PCR)	-	1
Total		50

Table 33. Reagents and volumes required to make a second round PCR master mix for one reaction.

49µl of second round PCR master mix and 1µl of DNA template was added to sterile PCR tubes. The reaction mixture was briefly mixed in a centrifuge then placed on ice. PCR reaction tubes were transferred to a thermal cycler for amplifications using the pre-set parameters (Table 32). A 3% agarose gel (1.8g of agarose) was prepared and used to separate LAM-PCR amplified products. 15µl of LAM PCR products was mixed with 3.5µl of loading buffer (Fisher Scientific Cat # BPE633-5) and loaded onto the gel. The gel was run at 50 V for approximately 1 hr.

2.2.13 454 sequencing

100-300ng of sample genomic DNA was subjected to deep parallel 454 pyrosequencing by Dr Manfred Schmidt (National Center for Tumor Diseases and German Cancer Research Center, Heidelberg, Germany) using the Roche GS FLX system.

2.2.14 RNA extraction, purification and cDNA synthesis for RT (reverse transcriptase) and QPCR (Real -Time Quantitative (Q) – PCR) analysis

RNA extraction for QPCR and RT PCR analysis.

Frozen liver tissue (-80ºC) was transferred to a mortar containing liquid nitrogen. Tissue was coarsely crushed using a pestle. Crushed tissue was extracted from the mortar using sterile forceps and transferred to a sterile 1.5ml centrifuge tube. 250µl of Tri Reagent was added to the sample and the tissue was homogenised using a sterile plastic rod. 500µl of Tri Reagent was added to the homogenised tissue. 200µl of chloroform was then added to the reaction mixture. The sample tube was mixed by inverting 5 times and left to incubate at room temperature for 15 minutes. Reaction mixtures were centrifuged at 12,000 g for 15 minutes at 4ºC in a Sigma centrifuge (model GK10). Following centrifugation the reaction mixture was separated into 3 layers representing RNA solution (clear aqueous phase), proteins (white phase) and DNA (pink phase) respectively. The clear upper fluid was transferred to a sterile 1.5ml centrifuge tube and 500µl of isopropanol (2-propan-1-ol) added. The reaction mixture was mixed by inverting tube several times and left to incubate at room temperature for 10 minutes. The reaction tubes were re-centrifuged for 10 minutes. The supernatant was removed using a pipette and the pellet re-suspended in 1ml of 75% ice cold ethanol to wash. The reaction mixture was vortexed followed by centrifugation at 12,000 g for 5 minutes (4ºC). This washing step was repeated once. The supernatant was removed and the tubes left to air dry for 15 minutes. The RNA pellet was then re-suspended in 50µl of DEPC water.

DNase I treatment

A DNaseI reaction mixture was prepared using the reagents and quantities listed in Table 34.

Reagent	Final concentration	Working volume x1 (µl)
RNA sample	-	50
Reaction buffer	1x	5
Amplification Grade *DNaseI*	1U/µL	5

Table 34. Preparation of a one reaction mixture for *DNase I* treatment.

DNase I was used to digest DNA present in RNA samples into oligo and mononucleotides prior to cDNA synthesis. The reaction mixture was left to incubate at room temperature for 15 minutes. 5µl of Stop *DNAse I* (50mM EDTA) solution was added to the reaction mixture and the solution mixed by brief centrifugation. Reaction tubes were transferred to a 70ºC heating block for 10 minutes. The reaction tubes were cooled in ice.

RNA purification for Real -Time Quantitative (Q) - PCR

Total RNA and mRNA samples were extracted and RNA was purified using the Absolute Total RNA & mRNA Purification Kit (Agilent technologies, Stratagene Cat # 400806). 1.75µl of β-ME was added to total RNA samples. 250µl of lysis buffer and 250µl of 70% ethanol was added to each reaction mixture and mixed by centrifugation 12,000 x g. The reaction mixture was transferred to an RNA binding spin cup and seated in a 2ml receptacle tube. The reaction mixture was mixed by centrifugation for 1 min. The flow through was discarded and 500µl of 1x low salt wash buffer added to the spin cup. The tube was re-centrifuged for 1 min and the flow through discarded. The addition of wash buffer followed by centrifugation and removal of flow through was repeated once with 500µl and again 300µl of wash buffer. The reaction mixture was re-centrifuged for 2 minutes. The spin cup was transferred to a sterile 1.5ml micro-centrifuge tube and 50µl of elution buffer added. The reaction mixture was incubated at room temperature for 2 min then re-centrifuged. The elution, incubation and centrifugation step was repeated once. Purified RNA samples were placed in ice and prepared for immediate cDNA synthesis.

2.2.15 CDNA synthesis for RT PCR

The absorbance values of purified RNA samples were found using the Nanodrop spectrophotometer. Second strand synthesis was performed in two steps RNA primer and cDNA synthesis using Superscript III First Strand Synthesis System (Invitrogen Cat # 18080-051). RNA primer master mix was prepared using the reagents and quantities listed in Table 37.

Reagent	1x working volume (µl)
5µg of total RNA	8
50µM Oligo $(dT)_{2o}$ primer	1
10mM dNTP mix	1
ddH_2O	10

Table 35. Preparation of a RNA primer master mix for one reaction mixture.

The reaction mixture was incubated in a 65ºC water bath for 5 minutes then placed in ice for 1 minute. A cDNA synthesis master mix was prepared using the reagents and quantities listed in Table 35 and placed on ice.

Reagent	1x working volume (µl)
10x RT buffer	2
25Mm $Mgcl_2$	4
0.1M DTT	2
40U/µl RNaseOUT	1
200U/µl Superscript III RT	1

Table 36. Preparation of a cDNA synthesis master mix for one reaction mixture.

10µl of this master mix was added to the RNA primers reaction mixture then mixed by centrifugation for 30 seconds. The reactions were terminated using heat inactivation by incubating tubes in a 85ºC heating block for 5 minutes then transferred to ice for 1minute. Reaction mixtures were briefly re-centrifuged followed by the addition RNaseH (1µl) to each tube. Reaction tubes were incubated in a 37ºc heating block for 20 minutes then stored at -20ºC until ready for use.

2.2.16 RT PCR amplification reagents

Primers were designed for mouse genomic sequences identified by sequencing. The expression of these genomic sequences was examined using RT PCR analysis.
A RT PCR master mix was prepared using the reagents and quantities listed in Table 37. 19µl of master mix was added to1µl of FIV infected cDNA samples (236ng/µl). Control reactions included gender matched normal livers and tumours without FIV insertions and blank negative controls.

Chemicals/Reagents	Final concentration	Volume x1 (µl)
ddH_2O	-	11
PCR buffer	1x	2
$Mgcl_2$	2.5mM	1
dNTP	0.4mM	2
Forward primer	Variable dependant on reaction	1
Reverse primer		1
Taq DNA polymerase	0.2U/µL	1
Total		19

Table 37. Preparation of RT PCR master mix for one reaction mixture.

2.2.17 cDNA synthesis for QPCR

cDNA was prepared from purified total RNA samples isolated from liver and gonad liver tissue. All RNA samples were quantified using the Nandrop spectrophotometer. For control experiments a standard curve was designed to optimise the PCR amplification. The starting amount of RNA was 1µg in a total volume of 10µl. Following this a two-fold dilution of the RNA samples was made using ddH_2O to yield a series of RNA templates with different concentrations. 6 dilutions were made per sample to yield a final sample containing 15.6ng in 10µl. The optimised input of RNA template was 125ng in 10µl of RNA. All reverse transcription reactions were performed on ice each RNA sample using the reagents and quantities listed in Table 38.

Reagents	Working volumes x1 (µl)
ddH_2O	3.2
10X RT Buffer	2
10X RT Random Primers	2
25X dNTP mix (100mM)	0.8
MultiScribe™, Reverse *Transcriptase*, 50 U/µL	1
RNase Inhibitor	1

Table 38. General reagents/chemicals used for CDNA synthesis of total RNA for QPCR reactions.

10µl of RNA sample was added to each reverse transcription reaction and mixed briefly centrifugation. The reaction mixtures were transferred to thermal cycler and reverse transcription performed using the parameters listed in Table 39.

Step	Temperature	Time
Primer extension	25ºC	10 min
cDNA synthesis	37ºC	120 min
Reaction termination	85ºC	5 min
Pause	4ºC	-

Table 39. PCR parameters used for cDNA synthesis using MultiScribe reaction kit.

2.2.18 Q-PCR for gene expression analysis and VCN analysis

A real time PCR reaction master mix was prepared on ice using the reagents and quantities listed in Table 40. The reaction mixtures were aliquoted into a 96 well plate. 2µl of cDNA was added to each reaction master mix. The amount of DNA required for gene expression analysis of FIV insertion genes was 125ng of cDNA. For VCN experiments 2µl (15.63ng/µl) of genomic DNA was added to each real time reaction master mix. The reaction plates were sealed with a MicroAmp® 96 Optical adhesive film (Applied Biosystems Cat # 4311971) to prevent evaporation & loss of samples during PCR reactions.

Reagent	Working volume x 1 (µl)
ddH_2O	7
2x TaqMan Universal PCR Master Mix	10
20x TaqMan gene expression assay	1

Table 40. Reagents/chemicals used for preparation of TaqMan PCR mastermix for amplifications.

The reaction mixtures were transferred to a 7900HT real time PCR thermal cycler (Applied Biosystems). PCR amplifications were run in the absolute quantification blank template format. The parameters for all amplification reactions are listed in Table 41.

Step	Temperature	Time	No. cycles
....	50ºC	2 minutes	1x
Denaturation	95ºC	10 minutes	...
Denaturation	95ºC	15 seconds	40 x
Annealing	60ºC	1 minute	
Extension	72ºC	90 seconds	

Table 41. PCR parameters used for TaqMan amplification reactions.

Validation experiments for PCR efficiency and optimal template concentration required were calculated quantitatively using a standard curve. An R^2 value of 1 indicated the ability to predict template amount with the Threshold Cycle (CT) value. Raw data for each reaction plate consisted of CT values. CT values were manually transferred from the thermal cycler to a CD for data analysis.

Q-PCR for gene expression analysis:

All reactions were performed in quadruplets and the reactions repeated on two occasions. All genes were normalised using the house keeping gene (*18sRNA*) and normal liver samples. A gender matched normal liver sample was used as the reference sample in each experiment. The relative expression level of each gene was manually calculated from CT values using the delta delta CT (ΔΔCT) method.

1. The average CT for each gene was subtracted from the average housekeeping CT value to produce a ΔCT value.
2. ΔCT values from each sample was subtracted from the reference sample to yield a ΔΔCT value (ΔΔCT = ΔCTsample – ΔCTnormal liver).
3. Relative gene expression values $2(^{-\text{log fold values}})$ were calculated using the following equation - $2^{-\Delta\Delta CT}$.

Statistical analysis including 95% confidence intervals, standard error of the mean (SEM) and student T testing was performed on CT values to validate gene expression data between normal and tumour samples.

Q-PCR for determination of vector copy number (VCN)

For VCN analysis raw CT values for vector specific sequences and housekeeping genes were converted into VCN values. A standard curve was obtained from a serial dilution of a tumour control sample with a known copy number of 1. A standard curve was constructed by plotting Log_{10} genome values (x axis) against CT values (Y axis).VCN values were calculated using a template excel spreadsheet containing all relevant calculations. The linear regression equation (y = mx+c) where m = slope and c = intercept was obtained from the standard curve. This equation was used to interpolate the VCN for all samples with unknown copy numbers and compare to tumour samples with known VCN (see Appendix 7 - Southern blotting).

1. For each sample the numbers of sequence copies were calculated by subtracting the average GAPDH CT from the intercept value (standard curve - GAPDH probe). This value was then divided by the slope value obtained from the standard curve.
2. Step 1 was repeated using values obtained with WPRE/EIAV psi data.
3. To calculate the copy number per haploid genome the total number of virus specific sequences detected were divided by the calculated total number of GAPDH.

$$\frac{\text{Copies of virus specific sequences (WPRE/EIAV psi)}}{\text{Copies of GAPDH}} = \text{Copy number per haploid genome}$$

4. Calculated haploid copy numbers were multiplied by 2 to obtain copy number per diploid genome (VCN).

2.2.19 Microarray analysis for gene expression of murine cellular genes retrieved by LAMPCR and 454 sequencing and known E2F targets

Preparation of RNA and collation of raw data for Microarray was conducted by Dr Manfred Schmidt (National Center for Tumor Diseases and German Cancer Research Center, Heidelberg, Germany). All microarray data was adjusted and classified according to individual p-values to create comprehensive lists of differentially expressed genes. This was done in collaboration Dr Rob Andrews at the Welcome Trust Sanger Centre, Cambridge, UK.

2.2.20 Bioinformatics – STRING analysis

Computational analysis was used to find genes related to genes found flanking the provirus insertion site. This enabled the analysis of the microarray data and to determine potential pathways associated with HCC. Suggested genes were entered into the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database (http://string-db.org/). This database contains a combination of known and predicted proteins. For analysis of protein-protein interaction raw network data is scored for all interaction partners with values listed in descending order. Each network is then categorised and clustered to reveal tightly connected functional partners and direct and physical associations.

2.2.21 Tissue culture

All tissue protocols were performed under a laminar class II cell culture hood (Heraeus). HepG2 cells were cultured in culture dishes (Fisher Scientific) containing 10ml of complete DMEM medium containing pen/strep and FBS.

To make 50ml of complete medium at 15% FBS the following reagents were added to a falcon tube: 42ml DMEM (Fisher, VX21885025), 7.5ml FBS (Fisher, VX10270106) and 495µl pen/strep (Fisher, VX15400054).

The culture dishes were transferred to a 37ºC incubator in a humidified atmosphere (5% CO_2). The cultures were allowed to grow until 70% confluent before proceeding to passaging steps. The culture medium (10ml) containing cells were transferred to a 50ml falcon tube. 10ml of sterile 1x PBS was used to wash the cells (by swirling) and the solution placed in a 50ml falcon tube (containing complete medium).

Confluent cells were detached from culture dishes using 1ml of 1x trypsin. The culture dishes were then incubated at 37°C for 5 minutes. 10ml of complete medium was added to each dish and gently mixed by pipetting up and down. The neutralised cell suspension was centrifuged at 1000rpm for 5 minutes. The clear supernatant formed was discarded. The pellet was re-suspended in 1ml of fresh (complete) medium and mixed by pipetting using a p1000 Gilson pipette. 1ml of cells suspension was added to 29ml of fresh (complete) medium and pipette up and down. 10ml of each cell suspension was transferred to a three sterile culture dishes and transferred to a 37°C incubator for growth.

2.2.22 Storage of cells derived from tissue culture

Cells were frozen in liquid nitrogen (-196°C) for long term storage or preparation for tissue extraction. Cells pellets obtained from centrifugation during passaging were re-suspended in freezing medium containing 1ml of culture medium & 10% DMSO The cell suspension solution was cooled at -80°C then transferred to a dewar containing liquid nitrogen.

2.2.23 DNA extraction from frozen tissue samples and cultured cells

All surfaces and equipment (forceps, pestle and mortar) were wiped own with 2% trigene and 70% ethanol. Frozen tissue samples were crushed to a fine powder in liquid nitrogen using a sterile pestle and mortar. Homogenised tissue samples were diluted with 2-3ml of extraction buffer and mixed by stirring with sterile pipette until a thick gluppy solution was generated. The reaction mixture was transferred to a sterile 50ml centrifuge tube using a pipette. Proteinase K at a concentration of 100µg/ml and RNAse was added to the reaction tubes and incubated at 55°C for 2 hours. An equal amount (2-3ml) of phenol was then added to the reaction mixture and mixed by gently inverting tube sideways until a white solution was formed. The reaction mixture was spun in a centrifuge at 6,000 rpm for 10 min at 4°C. The reaction mixture yielded 2 phases separating DNA and protein between the phenol and TE solutions. The top layer was DNA in aqueous solution. The aqueous DNA solution was transferred to a sterile centrifuge tube and an equal volume of chloroform was added. The solution was then mixed carefully by inversion. The reaction mixture was spun down at 6,000 rpm for 5 min at 4°C. The supernatant was

removed and transferred to a sterile centrifuge tube. 2-3ml of ice cold 100% ethanol was added to wash the reaction mixture and the tube gently mixed by tilting until a clear solution with DNA precipitate was formed. The solution containing DNA precipitate was spun down to form a pellet. The ethanol solution was poured off by tilting the tube in the opposite direction of the DNA pellet. DNA pellets were transferred to separate bijou tubes using a sterile pipette. The DNA pellets were washed in 70% ethanol to remove excess salt. The tubes were air dried in a sterile hood and the pellets re-suspended in 3ml of double distilled water. The reactions mixtures were left to dissolve at 4°C overnight.

2.2.24 Global methylation assay using Imprint® Methylated DNA quantification kit

DNA samples that were extracted from lentiviral (EIAV, FIV & HIV) infected liver and gonad tissues were quantified using a spectrophotometer. DNA samples were diluted with DNA binding solution (Sigma-Aldrich Cat # MDQ1) to yield a concentration of 50ng in 30µl. Standard control samples consisted of a negative blank and a positive DNA sample (uninfected and standard methylated DNA samples). The negative controls and methylated DNA (Sigma-Aldrich Cat # MDQ1) were also diluted in DNA binding solution. All dilutions were mixed by brief centrifugation at 12,000 x g and 30µl of each sample was added to an ELISA plate. To ensure that each well was coated the plate was gently tilted from side to side. The plate was covered with optical adhesive film (Applied Biosystems Cat # 4311971) and incubated at 37°C for 1hr. 10x wash buffer was thawed on ice. Following incubation 150µl of blocking buffer solution was added directly to each well to coat samples and reduce non-specific DNA binding. The plate wells were covered and incubated at 37°C for 30 minutes. The reaction mixtures were removed from each well by inverting plate. Each well was washed by adding 150µl of 1x wash buffer followed by inversion of the plate to remove the contents. This step was repeated 3 times. Methylated DNA capture involved the use of a capture antibody specific to methylated CpG dinucleotides. A 1X wash buffer was prepared in a sterile bottle using 11ml of 10x wash buffer (Sigma Aldrich Cat # MDQ1) and 99ml of ddH_2O. Capture antibody was diluted in a 1:1000 ratio using 1x wash buffer. 50µL of diluted Capture antibody was then added to each well. The plate was then covered and

incubated at room temperature for 1hr. The capture antibody was removed from each well by inverting the plate. Each well was washed 4 times with 150µl of 1x wash buffer. The detection antibody was diluted to a 1 in 1000 ratio using 1x wash buffer. 50µl of diluted detection antibody was added to each well. The plate was covered and incubated at room temperature for 30 minutes. The detection antibody was removed from each well and the reaction wells were washed 5 times with 150µl of 1x wash buffer. For the detection of methylated DNA 100µl of developing solution was added to each well. The plate was covered and incubated at room temperature away from light for 1-10 minutes. The reaction mixtures were monitored for colour change to ensure the formation of a blue solution. 50µl of stop solution was added to each well which yielded a yellow solution. The absorbance values correspond to the level of CpG methylation. Absorbance (Abs) values were measured at 450nm on a plate reader (Biotex Instruments). To calculate the relative methylation levels for each DNA sample. Firstly replicate absorbance values for all DNA and blank samples were averaged. The average blank value was then subtracted from the average absorbance values for each DNA sample to give value A. Next the average blank value was subtracted from the average absorbance value taken from the positive control sample to give value B. Value A was divided by value B and multiplied by 100. This calculated percentage value represents a global methylation level that is a percentage of the positive control DNA sample.

Calculation for global methylation: $$\frac{\text{(Abs 450 sample - Abs 450 Blank)}}{\text{(Abs 450 Methylation Control DNA - Abs 450 Blank)}} \times 100$$

2.2.25 Quantification of DNA methyltransferase activity using QPCR

Several gene expression assays were used to assess the expression of DNA methyltransferases in lentiviral infected samples. These results were compared to the expression of *Dnmt's* in uninfected cDNA samples. *18sRNA* was used as the housekeeping gene to normalise gene expression. The reagents for the real time PCR were prepared on ice. A single PCR master mix was prepared using the reagents and quantities listed in Table 40. 2µl of cDNA was then added to 18µl of PCR master mix in 96 well reaction plates. The raw data (CT) values were analysed via ΔΔCT analysis (see section 2.2.18).

2.2.26 Bisulphite modification, methyl sensitive PCR and DNA sequencing

Bisulphite modification, PCR and sequencing analysis was done in collaboration Dr Rob Andrews at the Welcome Trust Sanger Centre, Cambridge.

2.2.26 Brief on statistical methods used

All QPCR reactions were performed in quadruplicates for each sample. Each reaction was then repeated again (x4) on another occasion. An n value of 8 was used for T-test analysis. For microarray analysis fold changes in gene expression (logFC) were calculated and statistical analysis performed on raw data. For IPA analysis the Fisher's exact test was used to determine the statistical significance of genes enriched within a given GO category and validate genes overrepresented in data sets. P-values$<10^{-20}$ or lower were used to select highly significant overrepresentation after Benjamini Hotchberg correction.

3.1 Investigation of vector associated geneotoxicity in mice following *in utero* delivery of EIAV, FIV and HIV vectors.

3.1.1 Detection of viral vectors (EIAV, FIV and HIV) in infected mouse normal and tumour tissue

In 2005 the Gene Therapy Research group at Imperial College London under the supervision of Dr Michael Themis reported that 80% of haemophiliac and MF-1 mice treated with SMART EIAV based vectors *in utero* at day 16 gestation developed HCCs that were clonal in origin. Conversely, no HCCs were observed when the HIV vector HR'SIN-cPPT-S-FIX-W was used. The virus copy numbers (VCNs) found in these tumours varied between 1 and 10 for EIAV SMART vectors and in normal livers mean VCN's were 0.9 (+/-0.09) and 0.8 (+/-0.29) for the EIAV and HIV vectors, respectively. As insertional mutagenesis was suspected to be the cause of oncogenesis provirus integration sites were investigated. To do this LAM PCR was used to obtain 26 provirus integration sites from liver and gonad tissue. Of these, 20 were matched to RefSeq genes that were either cancer genes or cancer associated. This suggested that lentiviral transduction could have caused the oncogenesis in the fetally treated adult mice. During this investigation the WPRE was suspected also to be involved in tumour development. The was because the EIAV vector carries the woodchuck post translation regulatory element (WPRE) which contains the X gene which in the wild-type 140 amino acid form is known to be involved in HCC formation (Zufferey *et al*, 1999; Kingsman *et al*, 2005; Higashimoto *et al*, 2007; Schambach *et al*, 2007). In the EIAV and HIV vectors the sequences for a truncated form of the X gene was included but only in the EIAV vector is the start codon and X promoter (native form), that could allow X gene expression in contrast to the HIV vector sequences mutated to prevent X gene expression. Hence it was important that the role of the X gene was investigated as a possible factor leading to oncogenesis.

Following the original work carried out by the Themis group, a modified EIAV vector SMART2Δ and a feline immuno-deficiency virus FIV pLIONhAATGFP that has been shown to effectively transduce fetal mice (Condiotti *et al*, 2004) was used alongside the original SMART2 vector and the HR'SIN-cPPT-S-FIX-W HIV vector

that did not cause oncogenesis in fetal mice. This study was used to test the WPRE in EIAV and FIV and to determine whether the observed oncogenesis was restricted to EIAV vectors only. Mice treated with these vectors developed tumours at a frequency of 4/6 (SMART2 unmodified), 4/10 (SMART2 modified vector) and 4/8 (pLIONhAATGFP). The vectors used for this work are shown in Figure 13.

The research work described in this thesis began by investigating integration sites of the FIV vector in the mouse genome in order to identify common insertion sites (CIS) that could give clues on a particular gene which, if mutated, could be responsible for HCC following virus integration. The vectors and tissue samples obtained following *in utero* gene transfer to fetal mice were provided by Dr Themis following the genotoxicity studies described above (not previously published) and those previously described by the Themis research group (Themis *et al*, 2005).

3.1.1.1 Sequence comparison of vector sequences *in vivo*

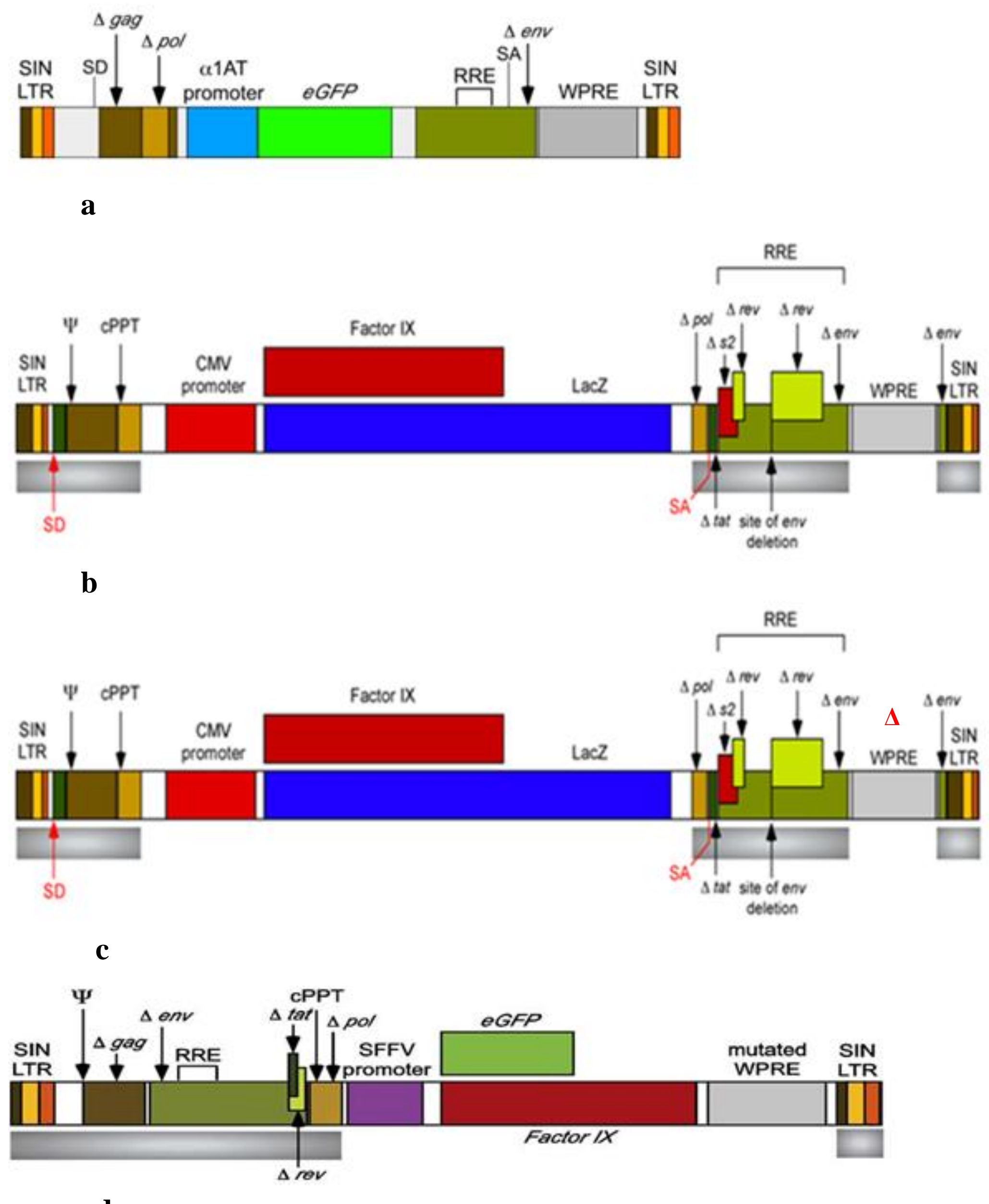


Figure 13. Schematic representation of the genomic organisation of vector constructs used to treat MF-1 mice *in utero*. Vector constructs used were pLIONhAATGFP, SMART2hFIX, SMART2Z, SMART2hFIX WΔ, SMART2ZWΔ and HR'SIN-cPPT-S-FIX-W respectively. All lentiviral constructs included SIN LTR sequences which contain a mutation in the U3 region of the 3 LTR. a) PLIONhAATGFP contains an antitrypsin (a1AT) promoter and enhanced green fluorescent protein (eGFP) reporter gene. b) All SMART2 vectors were driven by a CMV promoter. However each vector contained either a Factor IX (SMART2hFIX) or a LACZ transgene (SMART2Z). Modified SMART vectors (SMART2hFIX WΔ and SMART2ZWΔ) contained a mutated WPRE sequence (WΔ). d) The HR'SIN-cPPT-S-FIX-W construct used a SFFV (Spleen Focus-Forming Virus) to drive transcription. HIV vectors used in this study included either eGFP or Factor IX transgenes.

3.1.1.2 Comparison between VCN data obtained from Southern blotting and QPCR

Montini *et al* (2009) showed that VCN can contribute to vector mediated oncogenesis using a mouse model predisposed to oncogenesis with mutations effecting both p53 and Rb pathways.

VCN was established in EIAV, FIV and HIV infected liver tissues from the mice and shown in Table 42. Standards for VCNs were obtained from Southern blot analysis of clonal mouse tumours infected by EIAV vectors (see Appendix 7) carried out prior to the work presented here. These Southern blots showed clear bands of provirus insertions with each band representing a single proviral integration site. Based on these VCNs QPCR was used to establish VCN in infected tissues. QPCR failed to detect provirus insertions in uninfected tissues as expected but identified insertions in the known clonal tumour samples. QPCR results of animal samples are shown in Table 42.

Table 42. Vector copy number in infected livers and tumours

Vector	Mouse name	Gender	Age of mouse	Normal liver VCN (+/- SEM)	Tumour VCN (+/- SEM)
-	Uninfected	Male	578	Uninfected - 0 (0)	0
SMART2Z	Angus	Male	573	9.80 (0.62)	5.78 (0.42)
	Agnus	Female	487	18.69 (0.08)	8.61 (0.18)
	Archie	Male	531	2.07 (0.14)	2.09 (0.02)
	Adame	Female	644	4.23 (0.09)	2.89 (0.67)
SMART2ZWΔ	Gina	Male	127	4.81 (0.11)	8.05 (0.56)
	Gerty	Female	162	0.004 (0.01)	NI
	Golliath	Male	537	5.74 (0.00)	4.34 (0.22)
	Gladys	Female	369	2.61 (0.04)	6.47 (0.14)
	Gerry	Male	279	45.16 (0.09)	T(1) 86.34 (0.10) T(2) 202.13 (0.03)
	Gonzo	Male	640	0.01 (1.12)	Bone mass 0.07 (0.79)
SMART2 hFIX	Jonas	Male	348	26.38 (0.14)	43.99 (0.09)
SMART2lacZ LCMV	Monika	Male	229	NI	2.28 (0.34)

SMART2lacZ rabies	Rachel liver tumour	Female	192	NI	1.33 (0.25)
SMARTZ Ebola	Mike	Male	359	NI	1.46 (0.35)
pLIONhAATGFP	Flora	Female	715	Flora N -	(OC)0.03 (0.08)
	Frank	Male	433	0.47 (0.78)	T(1)2.12 (1.43) T(2)5.84 (0.26)
	Fanny	Female	484	1.53 (0.03)	T(1)2.30 (0.26) T(2)12.75 (0.05)
	Fiona	Female	273	14.61 (0.14)	T(1)3.28 (0.34) T(2)5.15 (0.37)
HR'SIN-cPPT-S-FIX-W	Topsy	Female	657	0.73 (0.40)	(adrenal mass)None
	Tilley liver	Female	694	2.70 (0.00)	liver mass 0.32 (0.09)
	Kostas liver	Male		0.19 (0.10)	None
	Bruce liver	Female		0.21 (0.04)	None
	Pierre liver	Male		0.21 (0.10)	None

Table 42. Animal tissues subjected to QPCR to determine VCN using primer/probes that recognised EIAV Packaging Signal (PSI) of the EIAV SMART vectors and the WPRE sequence present in the HIV and FIV vectors found inserted in mouse (MF-1 strain) genomic DNA. Values obtained from samples with known copy number found by Southern analysis were used to calculate VCNs in infected livers and tumours. VCNs varied widely regardless of the dose of vector used for infection (Appendix 1). Hence, it is difficult to control VCNs in treated mice. Interestingly, VCNs in HIV infected samples were all below 3 in contrast to those found in EIAV and FIV tissues. These low VCNs may be the reason why no tumours were identified in HIV treated mice. NI represents samples not investigated in this study due to the absence of tissues. Samples with results of ''none'' failed to develop liver tumours. LCMV = lymphocytic choriomeningitis virus envelope, Ebola = Ebola Zaire envelope, Rabies = rabies envelope. OC = Ovarian cyst.

The mean VCNs in normal mice treated with SMART2Z and SMART2ZΔ were found to be 8.7 (SEM – 3.7) and 11.7 (8.4), respectively. Student T-test analysis failed to find any significant differences between VCN values obtained from normal mice treated with SMART modified and unmodified vector constructs. The VCN values obtained from mouse tumour samples infected with SMART 2Z and SMART2ZΔ were 4 (SEM - 1.5) and 36.3 (SEM - 22.7), respectively. Again T test analysis failed to identify any significance between the VCN values obtained from tumour samples of these groups. Interestingly Gerry T1 and T2 samples yielded high VCN values of 86.397 (0.10) and 202.314 (0.03), respectively. These samples contributed to the large range of VCN values represented by the SMART2ZΔ group. VCN values obtained from SMART2 vectors showed that VCN varied greatly between mice that received the same dose of vector (Table 42). The VCN results taken from normal FIV infected mice ranged from 0.474 (SEM - 0.8) to 14.606 (SEM - 0.1) in normal samples. In comparison VCN values obtained from FIV treated tumour tissue ranged from 0.026 to 12.752. Similarities between the VCN data obtained from normal and tumour FIV infected samples yielded P values that were not significant (>0.05). However, this range of VCN values seen in FIV treated mice show once again the large variation in VCNs in mice receiving the same vector dose. HIV VCN results were significantly lower than those collected from FIV tumour samples ($p<0.05$). Mice infected with HR'SIN-cPPT-S-FIX-W had a mean VCN value of 0.573 (SEM - 0.31). The low VCN values found in HIV infected samples suggest the absence of tumour development in HIV infected mice may be due low vector copies. However, Flora ovarian cyst produced a VCN value of less than 1 (0.03). This suggests that other factors apart from VCN may play a role in oncogenesis.

We determined whether the VCN obtained from QPCR analysis correlated with the Southern blot VCNs previously obtained (Table 43). These results were then compared to the diploid VCN per genome values found using QPCR. The QPCR VCNs closely agreed with those found by Southern analysis of Angus T, Adame T1, Gina T, Rachel and Fiona T2 tumour samples and confirmed a wide variation of vector copies between infected tissues. Certain tumour samples that had broad bands following hybridisations on Southern blot membranes prevented accurate calculation of provirus copies. Examples of these include those found in Gerry T1 and T2 samples.

Table 43. Comparison between VCN data obtained from Southern blotting and QPCR.

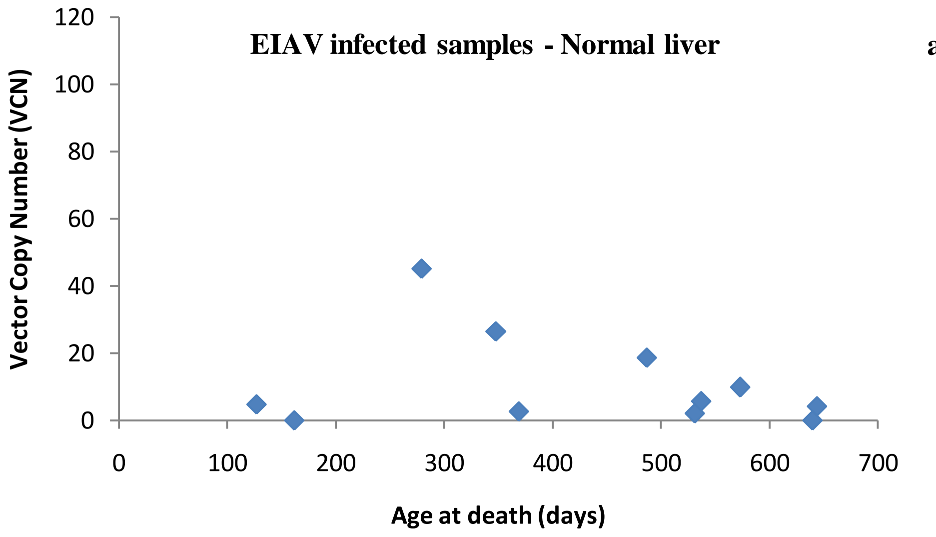
Tumour samples	QPCR VCN	Southern blot VCN
Uninfected	0	0
Angus T	5.78	9
Agnus T	8.61	10
Archie T	2.09	3
Adame T1	2.89	3
Gina T	8.05	8
Goliath T1	4.34	2+ smear
Gladys T	6.47	2 + smear
Gerry T1 Gerry T2	86.34 202.13	Thick smear difficult to identify specific bands 6 + thick smear
Gonzo BM	0.01	0
Jonas LT	43.99	5
Monika	2.28	1
Rachel	1.33	1
Mike	1.46	Smear
Flora Ovarian Cyst (FOC)	0.03	1
Frank T1 Frank T2	2.12 5.84	Thick smear difficult to detect individual

		bands 2
Fanny T1 Fanny T2	2.3 12.75	6 thick merging bands possibly more 5 thick merging bands possibly more
Fiona T1 Fiona T2	3.28 5.15	5 6

Provirus copies between samples examined using Southern blotting and QPCR were in close agreement. Where thick bands were found suggestive of two or more provirus insertions, estimation of VCN is given.

To determine any relationship between VCN and age of onset of oncogenesis, values for each parameter were plotted and examined. Figure 14 and 15 shows that no correlation between VCN and age of death exists following EIAV or FIV infection.

3.1.1.3 VCN measurements in EIAV infected mice



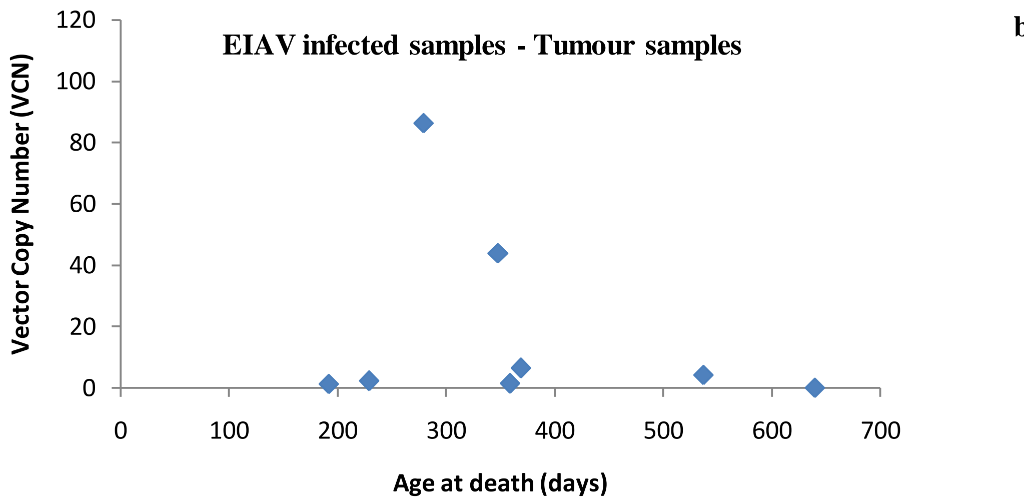
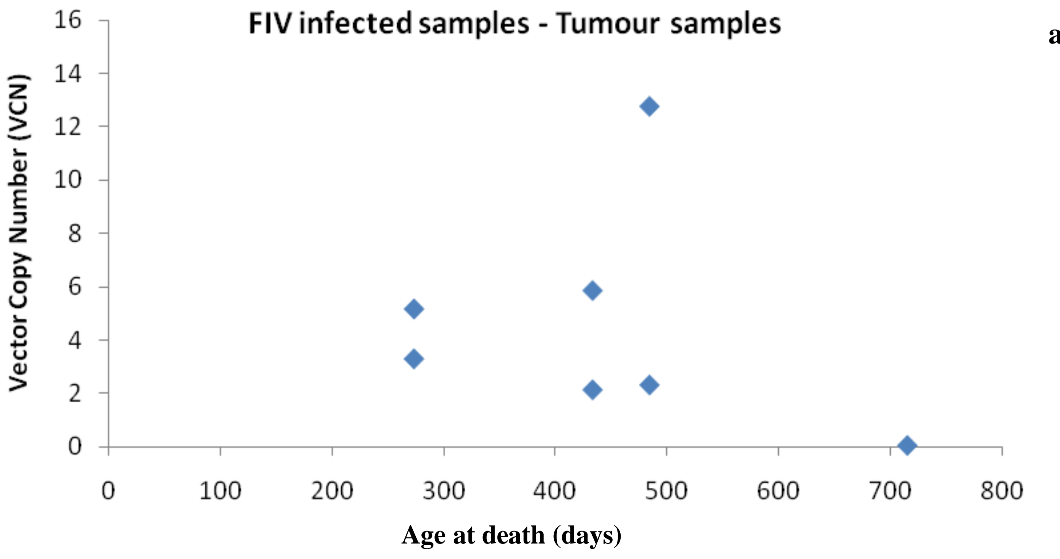


Figure 14. EIAV infected samples: Graphs **a.** and **b.** showing age at death versus VCN. Each dot represents a tissue sample derived from an EIAV infected mouse. The dose used for infections was $1.2\text{x}10^7$ and the age of death varied between 127 and 644 days in treated mice. VCN did not correlate with age of death suggesting other factors may be associated with vector may mediate oncogenesis.

3.1.1.4 VCN measurements in FIV infected mice

Mice treated with FIV vectors (3 normal and 7 tumours) were also examined to look for correlation between age of death and VCN. Again no correlation was found between age of death and VCN in FIV treated mice (Figure 15).



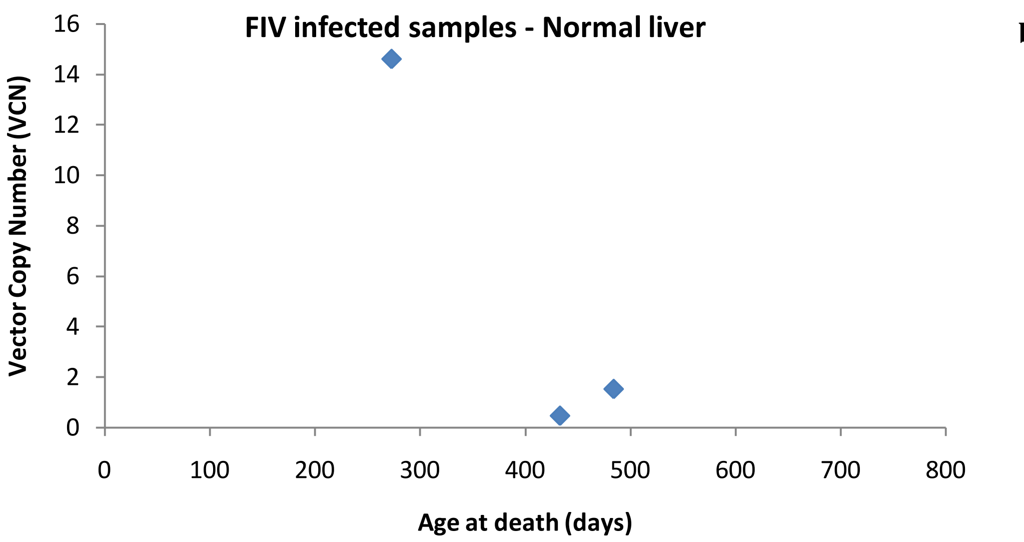


Figure 15. FIV infected samples: Graphs **a.** and **b.** of age at death versus VCN. Each dot represents tissue from an infected mouse. The dose used for infections was $1.0x10^7$. Ages at death varied between 127 and 644 days in treated mice and VCN did not correlate with age of death.

3.2.2 Retrieval of proviral insertion sites and flanking genomic sequences by LAM PCR and BLAST analysis

Previous gene therapy trials revealed that vector integration may produce unwanted side effects including clonal dominance and oncogenesis (Hacein-Bey-Abina, 2003; Ott *et al*, 2006; Schmidt, *et al*, 2007).

LAM PCR is able to detect vector genome junctions in a sample and retrieve integrations from clones even at the single cell level (Smidth *et al*, 2002; 2007; 2009). This has allowed for examination of integration site distribution and to study side effects of vector integration in the host genome. Several studies have utilised this method to retrieve retroviral integration sites (Schmidt *et al*, 2003, Themis *et al*, 2005; Hayakawa, *et al*, 2009).

In the previous study conducted by Themis *et al* (2005) LAM PCR analysis was used to retrieve 26 unique lentiviral insertion sites but failed to identify any CIS in tumour samples (n = 8) treated with SMART2 vectors. In this chapter vector integration sites were retrieved from FIV infected tissue samples to try to identify insertions that could be associated with oncogenesis. LAM PCR was performed on six liver tumours, one normal liver and one Ovarian Cyst sample. Unique integration sites were identified representing clonal insertions. In addition, 2 insertion sites were retrieved from non-clonal tissue. All LAM PCR products were run on agarose gels (Figure 17) and each band represents a single provirus insertion. Some normal samples gave the appearance of smears that are representative of multiple non-clonal insertions. LAM PCR did not detect proviral integration sites in uninfected control samples (Figure 17). Proviral insertions sites were also not identified in blank control samples (Lanes 1, 2 and 8). In order to reveal the identity of mouse genomic sequences LAM PCR products were cloned and subjected to DNA sequencing using an automated capillary DNA sequencer (Dundee DNA Sequencing Services, UK). Integration sites were visualised as chromatogram sequences (Figure 18).

Basic Local Alignment Search Tool (BLAST) is a program that allows researchers to compare protein or nucleotide queries with bioinformatic databases for sequence similarities (Altschul *et al*, 1997; McGinnis and Madden, 2004). Each LAM PCR product was exported into the BLAST program to retrieve sequences which shared a close homology with mouse genomic sequences found flanking the provirus.

The chromosome positions and names of well matched query sequences (>95%) were retrieved from the database. BLAST software was also used to annotate nucleotide sequences representing vector integration sites. Sequences annotated included the linker cassette, FIV LTR and pDRIVE cloning vector. A full list of known nucleotide sequences representing LAM PCR proviral insertions sites are provided in the Appendix 8.

LAM PCR revealed several unique insertion sites.

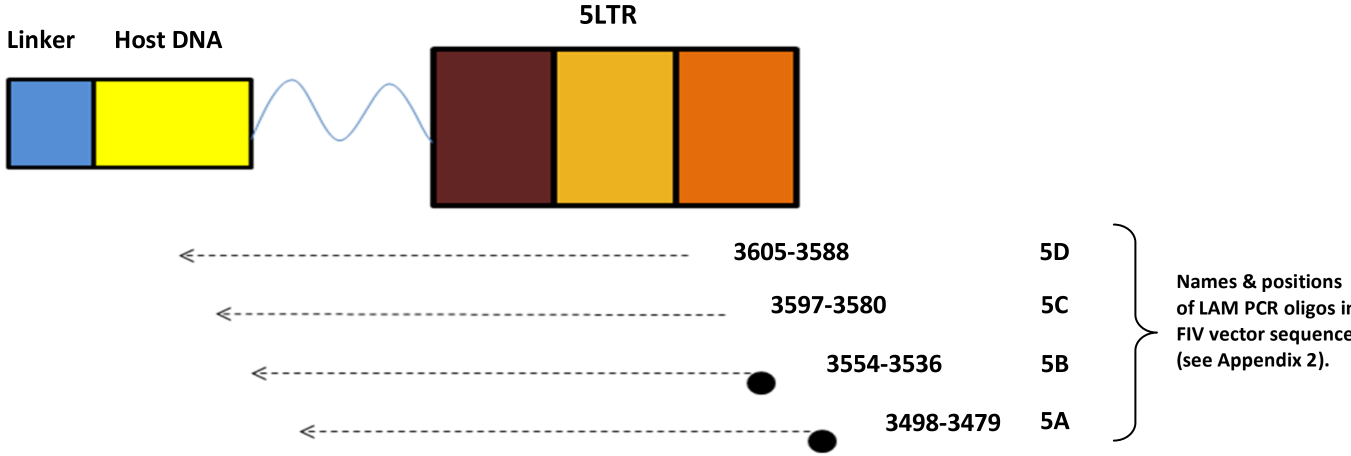


Figure 16. Schematic image of FIV vector and oligo sequences used in LAM PCR analysis. Oligo sequences were designed to hybridise to FIV LTR sequences. Biotinylated primers (5A and 5B) were used to produce biotin labelled linear PCR products containing the FIV vector and flanking genomic sequences. These oligos were used to generate a linear product containing viral LTR sequences and flanking mouse genomic sequences. 5A and 5B oligos were positioned at 3498-3479bp and 3554-3536bp, respectively. First and second round PCR steps used linker and vector specific (5C and 5D) oligo sequences. 5C and 5D oligo sequences hybridised to sequences located at 3597-3580bp and 3605-3588bp in the pLIONhAATGFP vector construct.

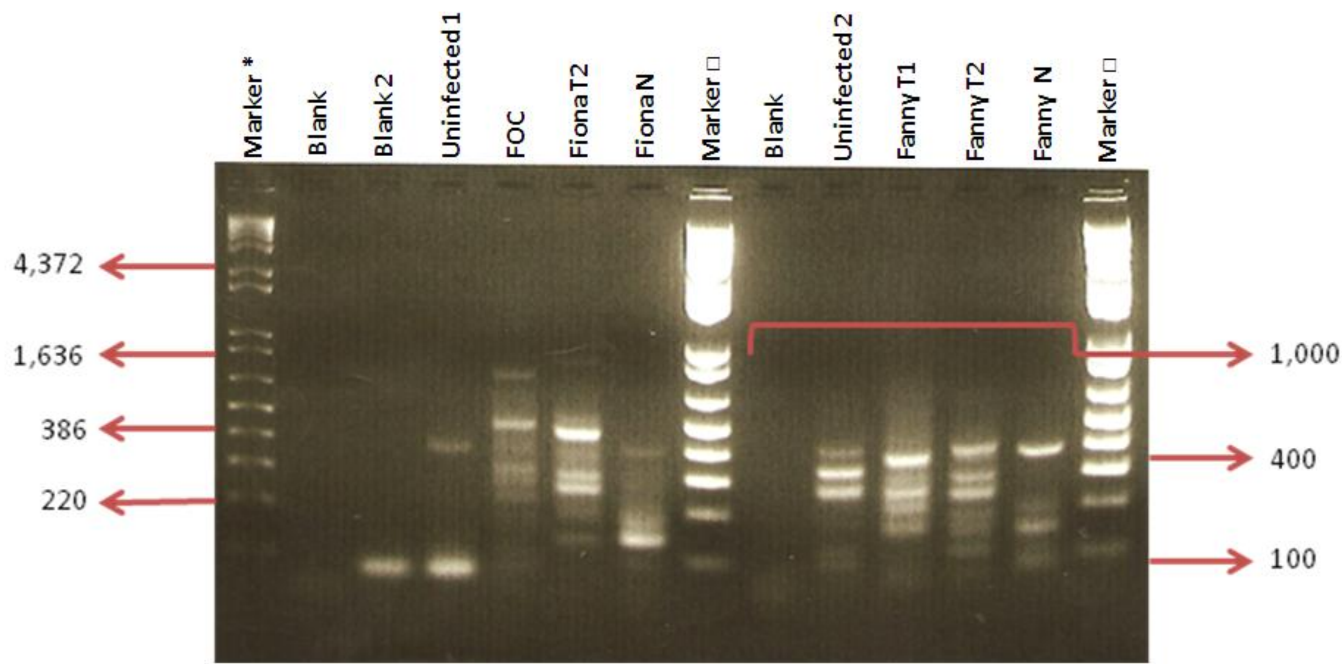


Figure 17. Proviral insertion sites retrieved from tumours: integrated provirus bands from each DNA sample was compared to two standard 1kb markers (* & □). LAM PCR products were cloned into the pDRIVE plasmid and sequenced using automated capillary sequencing methodology.

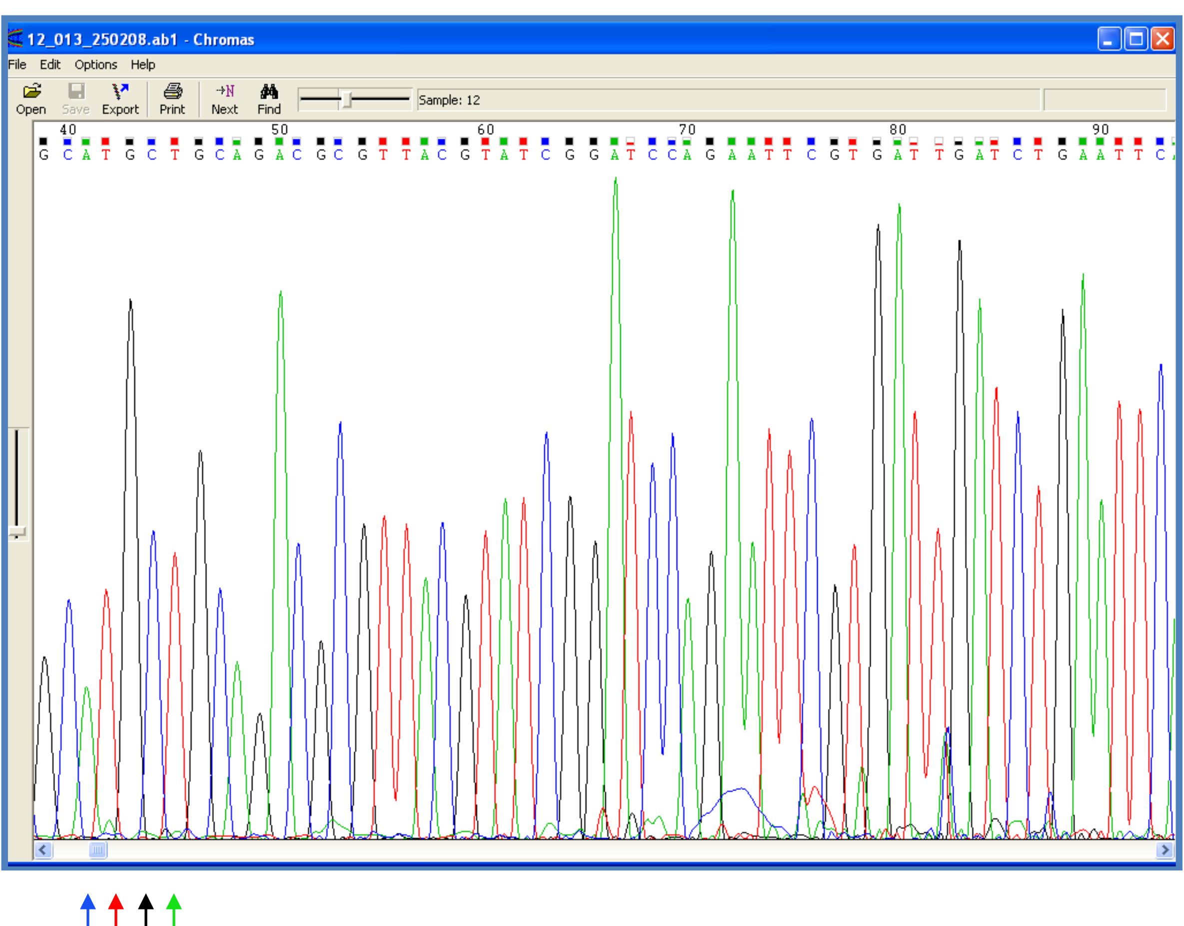


Figure 18. Schematic diagram of a chromatogram sequence representing a LAM PCR retrieved integration site from an FIV vector. Coloured peaks represent fluorescently labelled nucleotide bases. Blue, red, black and green peaks represent cytosine (c), thymine (t), guanine and adenosine (a) nucleotides, respectively. Chromatagram files containing mouse genomic sequences were exported to a FASTA format and aligned to the mouse genome using the BLAST (Basic local alignment search tool) database http://blast.ncbi.nlm.nih.gov/Blast.cgi (accessed online 2008).

3.2.3 Analysis of the position and role of EIAV and FIV insertion genes - gene ontology

The gene ontology (GO) of all provirus insertions was also identified using mouse genome databases including the MGI Mouse database http://www.informatics.jax.org/genes.shtml.

EIAV integration sites in tumours derived from EIAV infected mice were also retrieved using 454 sequencing. 454 sequencing is a highly sensitive method that generates simultaneous sequencing data ('reads') (Margulies *et al*, 2005).

25 FIV and 16 EIAV sequences were retrieved by LAM PCR and 454 sequencing. These sequences were examined to identify insertion sites within 100kb of known RefSeq genes.

Genomic sequences obtained from BLAST analysis were imported into the RTCGD Database (http://rtcgd.abcc.ncifcrf.gov/) (Table 44). This database describes the genomic position and orientation of several retroviral integration sites (RIS) that have been cloned from mouse tumours (Akagi *et al*, 2004). A 'hit' was used to describe a gene that had been previously identified and associated with oncogenesis in mouse tumour models. Common insertion sites (CIS) describes genomic regions with insertions occurring in more than one tumour and more likely to encode a disease causing gene. This study recovered several unique insertion sites in EIAV and FIV infected tumour DNA samples. Analysis of these insertions suggests that the HCCs are heterogeneous nature as no CISs were found (Table 44). This has previously been described for solid tumours where oncogenesis is believed driven by the activation of several pathways (Avila *et al*, 2006). Interestingly, 67% and 33% of FIV tumour insertions were positioned in the positive and negative orientation, respectively, to the transcriptional direction of the inserted gene (Table 44).

The distribution of proviral integration sites was also examined to identify any preferences for integration in the mouse genome. A small proportion of insertions found in this study have been described as RIS insertions in the RTCGD Database. 4 integrations (*Smg6*, *Tmem68*, *Bre* and *Pah*) from EIAV associated tumours were also RIS hits (Table 44). These insertion genes were obtained from Jonas and Adam tumour DNAs. RIS 'hits' were also identified in Frank T2 (*Coro7*), Fiona T2 (*Pscd3*) and Fanny T1 (NFI/B) tumour samples (Table 44). The RTCGD database also showed 'hits' for the *Smg6 (9)*, *Tmem68*, (1) *Bre* (2) and *Pah* (1) genes,

respectively. Interestingly an RIS 'hit' was also identified for *Clecd* a provirus insertion site isolated from Frank normal hepatic tissue. Although no CIS's were identified in this study, of the 24 insertions found 12.5 % and 16.5% were identified to be involved in signalling and cell cycle/cell division processes respectively. Interestingly 58% of FIV insertions are known to play a role in binding functions and mechanisms. The failure to identify CIS may be due to small number of provirus insertions studied. High sequence counts retrieved from the EIAV study found 7 CIS's. RefSeq genes and possible hotspots for provirus integration in Adame tumour tissue included *Pah*, LOC382044 and *Acvr2a*. 2 and 3 hotspots were retrieved from Gerry and Jonas tumour samples respectively (Table 44).

Table 44. Vector integration sites in FIV and EIAV infected tumours:

Mouse Number (Tumour)	Chr	Nearest' RefSeq Gene	ORI vector /gene	Identity	Molecular Function	Ontology and Biological Process	RTCGD CIS	RTCGD RIS	Relationship to Cancer
EIAV									
SMART 2ZWD Jonas LT	11	Int 10 of *Smg6*	+	Telomerase subunit EST 1A	DNA binding	Telomerase-dependent telomere maintenance	Yes	9 hits	Associated with cancer
	5	Int 2 of *Bre*	+	Brain and reproductive organ-expressed protein	Tumour necrosis factor receptor binding	DNA repair, apoptosis, response to DNA damage	None	2 hits	Oncogene
	3	5' of *Slc7a12*	+	solute carrier family 7 (cationic amino acid)	Amino acid transporter activity	Amino acid transport; Transport	No; 2 other Slcs are CIS)	None	None
	12	Int 1 of *Mark3*	+	MAP microtubule affinity-regulating kinase 3	ATP binding; Protein serine/threonine kinase activity; Transferase activity	Protein amino acid phosphorylation	None	None	oncogene; Wnt-signaling activator
	10	Int 1 of *Katna1*	+	Katanin p60 (ATPase-containing) subunit A1	Hydrolase, ATPase and Nucleoside-triphosphatase activity; Nucleotide binding	Cytokinesis; Protein localization; Mitosis; Cell cycle	None	None	choriocarcinoma-related gene
SMART2ZWD Jonas ST	4	Int1 of *Tmem68*	-	Transmembrane protein 68	Acyltransferase activity; Transferase activity	Phospholipid biosynthesis	No; 5 other Tmems are CIS	1 hit	Associated with oncogenesis
	3	Int 1 of *Foxo1*	+	Forkhead box O1a	Transcription factor activity; DNA binding	Insulin receptor signaling pathway; Regulation of proliferation/transcription	No; Foxo3a; 1 hit	None	TSG and oncogene associated
	9	Int 18 of *Smarce1*	+	SWI/SNF related matrix associated	DNA binding; Chromatin binding	Regulation of transcription, chromatin and structure	None	None	oncogene Brg1 associated factor
	8	3' of *Irs2*	-	insulin receptor substrate 2	Receptor activity, protein kinase binding	Cell proliferation, insulin receptor signaling, cell migration	No; Irs4; 1 hit	None	Ret oncogene co-factor
	19	3' of *Aldh1a7*	-	Aldehyde dehydrogenase family 1 subfamily A7	Aldehyde dehydrogenase (NAD) activity; Oxidoreductase activity	Retinal metabolism; Metabolism	No; Aldh1b1 1 hit	None	TSG

Mouse Number (Tumour)	Chr	Nearest' RefSeq Gene	ORI vector /gene	Identity	Molecular Function	Ontology and Biological Process	RTCGD CIS	RTCGD RIS	Relationship to Cancer
EIAV									
	9	3' of *Oxsr1*	-	Oxidative-stress responsive 1	ATP binding; Protein serine/threonine kinase activity; Transferase activity	Protein amino acid phosphorylation	None	None	None
SMART 2ZWD Gerry T1	6	Int 1 of *Setd5*	+	SET domain containing 5	None	None	None	None	None
	5	5' of *Mrpl1*	-	Mitochondrial ribosomal protein 1	RNA binding	RNA processing, translation	None	None	None
SMART 2Z Adame T1	10	Int 2 of *Pah*	+	Phenylalanine hydroxylase	Amino acid binding, metal ion binding, monooxygenase and oxidoreductase activity	Metabolic and oxidation reduction activity	None	1 hit	Reduced expression in cancer
	8	Int 2 of LOC382044	-	Predicted gene 5158	None	None	None	None	None
	2	5' of *Acvr2a*	+	Activin receptor IIa	Activin receptor activity, ATP, growth factor, metal ion, nucleotide binding, serine/threonine kinase activity and receptor activity	Activin receptor regulation of BMP signalling and nitric-oxide synthase activity, development, protein/amino acid phosphorylation, cell proliferation; serine/threoinine kinase signalling receptor	None	None	Associated with BMP in breast cancer

Mouse Number (Tumour)	Chr	Nearest' RefSeq Gene	ORI vector /gene	Identity	Molecular Function	Ontology and Biological Process	RTCGD CIS	RTCGD RIS	Relationship to Cancer
FIV									
pLIONhAATGFP Fanny T1	6	Int 10 of *Sgce*	+	Sarcoglycan epsilon	Calcim ion and protein binding; membrane stabilization	None	None	None	Associated with HCC.
	6	3' of *Peg10*	+	Paternally expressed 10	DNA and metal ion binding	Differentiation, placental development and apoptosis;	None	None	Associated with HCC
	6	5' of *Casd1*	-	CAS1 domain containing 1	None	None	None	None	None
	13	5' of *Ankrd32*	+	Ankyrin repeat domain 32	Protein binding	None	None	None	None
	4	Int 8 of *NfI/B*	+	Nuclear factor I/B	DNA binding; transcription activator protein; transcription factor activity;	Development, regulation of cell proliferation and transcription	None	3 hits	Oncogene; Blm TSG associated
	19	17288 bp at 5' side: *Pdcd1lg2*	+	Programmed cell death 1 ligand 2	protein binding, receptor activity	regulation of T-cell proliferation	None	None	None
	18	27028 bp at 5' side: *CX39/Gjd4*	-	Connexin 39/Gap junction protein delta 4	Gap junction channel activity	Cell communication	None	None	None
	18	7093 bp at 3' side: Similar to Cyclin fold protein 1- humans	-	Cyclin fold protein 1	Cyclin dependant protein kinase regulator activity; protein kinase binding	Cell cycle, cell division, Wnt signaling pathway	None	None	None
pLIONhAATGFP Fiona T1	3	Int 6 of *Pglyrp4*	-	Peptidoglycan recognition protein 4	N-acetylmuramoyl-L-alanine amidase activity; protein binding	Immune response; petidoglycan catabolic process	None	None	None
	1	5' of *Terf1*	-	Telomeric repeat binding factor 1	DNA binding, protein binding, telomeric DNA binding	cell cycle, division, mitosis, regulation of transcription, telomere maintenance	None	None	Associated with hepatocarcinogenesis
	5	5' of *Mob1*	+	Mps One Binder kinase activator-like 1A	kinase activity; protein, metal ion binding	protein modification process	None	None	Activates LATS1 TSG

	11	LOC627244 5' of Cob1	+	Cordon Bleu	Protein binding	Neural tube closure	None	None	Associated with microsatellite instability in colorectal tumours
		3'*Stk16*	+	Serine/threonine protein kinase	Kinase activity, DNA binding, ATP binding	Control of growth/development	None	None	Putative oncogene
	1	3' of *Tram1*	-	Translocating chain-associating protein 1	Protein targeting	Intracellular trafficking and secretion	None	None	None
pLIONhAATGFP Fiona T2	5	Int 1 of *Pscd3*	+	Pleckstrin homology, Sec7 and coiled-coil domains 3	ARF guanyl-nucleotide exchange factor activity	regulation of ARF protein signal transduction	None	2 hits	Associated with suppression of metastasis
	5	5' of *Usp42*	+	Ubiquitin specific peptide 4	ubiquitin cycle	None	None	None	None
	5	5' of *Slc16a4*	+	Solute carrier family 16, member 4	None	None	None	None	None
	5	3' of *Rbm15*	+	RNA binding motif protein 15	nucleic acid binding	Positive regulation of Notch receptor	None	None	Regulator of cMyc oncogene
pLIONhAATGFP Flora Ovarian Cyst	14	3' of *Sacs*	+	Sacsin	ATP binding, heat shock protein binding and lipoprotein receptor binding	None	None	None	None
	14	5' of *Tnfrsf19*	+	Tumour necrosis factor receptor superfamily, member 19	Receptor activity	Hair follicle development	None	None	Associated with gastric carcinogenesis
	13	5' of *Prl2c3*, 3' of LOC226089	+	Prolactin family 2, subfamily c, member 3	Growth factor activity; hormone activity;	Positive regulation of cell proliferation	None	None	Associated with breast cancer
pLIONhAATGFP Frank N	6	1053bp at 3'side of *Clec2d*		C-type lectin domain family 2 member d	Natural killer cell lectin-like receptor binding, protein binding, transmembrane receptor activity, sugar binding	Cellular defense response, negative regulation of osteoclast differentiation, protection from natural killer cell mediated cytoxcity, signal transduction.	None	1 hit	None
	6	93876bp at 5'side: *Clec2e*		C-type lectin domain family 2 member e	Sugar binding, receptor activity.	Signal transduction	None	None	None

pLIONhAATGFP Frank T1	12	Int 6 of *Nek9*	-	NIMA(never in mitosis gene A)-related expressed kinase 9	Metal ion binding, nucleotide binding, serine/threonine kinase activity and transferase activity	Cell cycle; cell division; mitosis; protein amino acid phosphorylation	None	None	Nek family involved in cancer
	3	Int 6 of *Pglyrp4*	-	Peptidoglycan recognition protein 4	N-acetylmuramoyl-L-alanine amidase activity; protein binding	Immune response; petidoglycan catabolic process	None	None	None
pLIONhAATGFP Frank T2	16	Int 9 of *Coro7*	+	Coronin 7	ND	ND	None	3 hits	None

Table 44. Vector integration analysis: a list of integration sites were mapped to the nearest RefSeq genes using the MGI database. The gene ontology of mouse chromosomes and genes identified by BLAST analysis were used to provide the molecular functions and biological processes listed in mouse genome databases. These generated GO terms for molecular function including binding, receptor activity and transferase activity and biological processes such as cell proliferation. The functional relationship of these genes to cancer was found using the RTCGD (retroviral tagged cancer gene database) and literature searches. Ori = orientation of vector into the host genome. Chr = chromosome identity.

4.1 Provirus integration can influence gene expression.

Provirus insertion has been shown to influence normal transcription of cellular genes through the use of potent regulatory regions (Coffin *et al*, 1997; Montini *et al*, 2006; Nienhuis, Dunbar and Sorrentino, 2006; Pajer *et al*, 2006; Coffin and Jern, 2008). The integration of RV DNA into the host-cell genome may result in insertional mutagenesis as discussed in chapter 1. The integrated provirus may affect adjacent genes through promoter insertion or enhancer (Uren *et al*, 2005; Martin-Hernández *et al*, 2006).

Gene expression changes are often a direct result of LTR insertions found upstream of genes in the antisense orientation and downstream in the sense direction (Coffin *et al*, 1997; Boonyaratanakornkit *et al*, 2004 Coffin and Jern, 2008). The danger is that interaction between the integrated RV LTR sequence and cellular transcription factors may result in activation of adjacent cellular proto-oncogene's and oncogenes (Gonda *et al*, 1987; Coffin *et al*, 1997; Montini *et al, 2006;* Coffin and Jern, 2008; Bosticardo *et al* 2009). Several proto-oncogenes encode transcription factors that are deregulated by mutation or elevated in human tumours (Degado *et al*, 2006).

Girad *et al* (1996) reported increased levels of *Notch1* in 52% of tumours as a result RV transduction in transgenic mice. This study showed that provirus insertion upstream of the *Notch1* oncogene was linked to elevated levels of full length and truncated *Notch1* proteins (Girad *et al*, 1996). Martin-Hernández *et al*, 2001 also demonstrated LTR mediated activation of N-ras and induced lymphoma development following provirus integration into the N-ras/unr locus.

4.1.1 Provirus and splicing mechanisms: the effect on gene expression

Other studies have shown that proviral insertions can knock out or truncate a gene (Harbers *et al*, 1984; Pajer *et al*, 2006; Voigt *et al*, 2008). Splicing is a cellular mechanism used to assemble genetic molecules and is crucial to the synthesis of genes and proteins in eukaryotes. RNA splicing describes the process by which introns are removed and exons joined to yield a primary transcript (Norton, 1994; Black, 2003; Matlin, Clarke and Smith, 2005). Several factors can contribute to alternative splicing including exon skipping, the presence of alternative donor sites and intron retention (Fackenthal and Godley, 2008). Introduction of provirus insertions can also induce genomic rearrangements through splicing and recombination mechanisms. This process is facilitated by the inclusion of viral splice donor and acceptor sites within the vector construct (Raineri and Senn, 1992; Brandt *et al*, 2005). The production of chimeric transcripts often results in altered expression of host genes. Gonda *et al* (1987) showed that RV integration can induce high levels of aberrant transcripts derived from the *C-myb* gene. A viral splice acceptor site was used to generate an alternative RNA product which encodes a truncated *Myb* protein and may contribute the oncogeneity of *Myb* in these tumours (Gonda *et al*, 1987). Overexpression of the *cKI-ras* protooncogene was observed in murine bone marrow derived cells following RV promoter insertion (Trusko *et al*, 1989). Trusko *et al* (1989) reported the production of an unexpected splicing event caused by the presence of a cryptic splice site positioned at the junction between and cellular DNA sequences.

4.1.2 Target site selection and gene expression patterns

The choice of RV used in transduction protocols is also known to influence the pattern and levels of transcription in host cells (Hu *et al*, 2008). De Palma *et al* (2005) used RV promoter traps to study integration site selection in primary haemopoetic cells. Results showed that MLV based vectors trapped cellular promoters more efficiently than HIV vectors. Previous observations have shown MLV based vectors to associate with promoter regions and HIV based vectors to associate with actively transcribed regions of the host genome (Schröder *et al*, 2002; Mitchell *et al*, 2004; Nowrouzi *et al*, 2006).

Provirus insertion may also be influenced by the nuclear architecture of chromatin (Albanese *et al*, 2008) and is also associated with epigenetic modifications (Wang *et al*, 2007). Gene expression is influenced by chromatin structure and recruitment of co factors (Delgado and Leon, 2006). Chromatin can be subdivided into open euchromatin and compact heterochromatin structures. The latter chromatin structure prevents interaction with transcription factors and disrupts transcription mechanisms (Delgado and Leon, 2006). An early study conducted by Conklin and Groudine reported varied integration patterns between provirus structures and adjacent host chromatin (Conklin and Groudine, 1986). However, this study demonstrated localised integration into euchromatin, exclusion of provirus insertions into heterochromatin regions and silencing of HIV-1 proviruses in the nuclear periphery. This suggests that integration sites are important to the expression of proviral and host cellular sequences (Dieudonné *et al*, 2009).

Proviral sequences may exert their effects on genes located away from the integration sites. Singhal *et al* (2011) reported long distance effects of retroviral vectors in mouse embryonic fibroblasts. This study showed a direct relationship between proviral insertions and aberrant gene expression in several mutant clones. Results showed that increased levels of RIPK1 and NF-κB expression were linked to provirus insertion (Singhal *et al*, 2011).

In some cases provirus insertion may have little effect on neighbouring genes. Recchia *et al* (2006) studied the impact of retroviral integration on gene expression in transplanted T cells. T-lymphocytes were transduced with retroviral vectors carrying the HSV-TK (Herpes Simplex Virus –Timidine Kinase) suicide gene and the truncated nerve growth factor receptor (ΔLNGFR).This study measured the

effect of gene expression in T cell clones and found that 1/5 of target genes located in the vicinity of promoter regions were upregulated. These target genes were shown to be expressed at low levels in T cells. Interestingly this study suggests that the upregulation of target genes was not influenced by the distance and orientation of the integrated vector (Recchia *et al*, 2006). A long term follow up study of *ex-vivo* transduced patients found that only around 2% of the >16,000 genes measured showed changes in expression (Recchia *et al* 2006). Furthermore changes in gene expression failed to effect normal T cell function (Recchia *et al,* 2006). Another study by Cassani *et al* (2009) performed quantitative transcript analysis on T cell clones taken from ADA-SCID patients. Microarray analysis was performed on 120 genes within a 200kb radius from integrated RV sites and results showed modest dysregulation of 5.8% genes in 18.6% of T cell clones when compared to controls.

4.2 Proviral insertions and the construction of cancer gene pathways

Proviral tagging has been used to retrieve cancer genes in oncogene bearing and transgenic mouse models (Jonkers *et al*, 1997; Sourvinos *et al*, 1999; Shi *et al*, 2002; Sivasubu *et al*, 2007; Kool and Berns, 2009). Mutational profiling of tumours and RV/LV integrations from genotoxicity models and insertional mutagenesis screens has revealed hundreds of genes which may contribute to oncogenesis. Cancer is indeed a complex disease that can arise from multiple spontaneous and or inherited mutations that result in dramatic changes in gene expression patterns and protein function (Thorgeiresson, Lee and Grimsham, 2006). It was recently reported that 384 human genes may have a causal role cancer in cancer development. Furthermore, cancer associated genes compromise almost 2% of the human genome (Santarius *et al*, 2010). In evaluating the relationship between RV and LV provirus insertions scientists can build potential gene pathways involved in signalling and cancer development (Callahan and Smith, 2000; Kool and Berns, 2009).

4.2.1 Hepatocellular carcinoma (HCC)

HCC is the fifth most common cancer worldwide and the third most common cause of cancer related deaths globally (Marrero and Marrero, 2007; Vinciguerra and Foti, 2008; Olsen, Brown and Siegel, 2010). The development of HCC is associated with various underlying liver diseases including autoimmune hepatitis, hepatitis virus infection, mycotoxins, hemochromatosis and cirrhosis (Llovet and Bruix, 2008). Studies have also reported aflotoxin B1 ingestion as a major risk factor in HCC development within sub-Saharan Africa and Asia but not in Western countries (Clavien and Breitenstein, 2009). Other studies have shown HCV infection to be a major pre-determinant - accountable for 70% of HCC cases in Japan (Clavien and Breitenstein, 2009). Several genetic mutations exist in HCC patients including that of p53 (Tannapfel and Wittekind, 2002). Disruptive mechanisms which contribute to these mutations include epigenetic events activation of oncogenes, inactivation of tumour suppressor genes and or loss of function mutations. Further work is needed to improve understanding of the molecular pathogenesis which contribute to HCC and identify key targets/markers for therapeutic treatment (Thorgeirsson, Lee and Grisham, 2006). Despite these multifactorial challenges recent experimental studies in animal and human models have been able to identify some mechanisms which are frequently disrupted in HCC and link common genetic mutations. Transcriptional analysis of HCC has identified that this disease is associated with several pathways including immune responses and protein deregulation (Thorgeirsson and Grimsham, 2006). Genome wide analysis performed by Lee and Thorgeiresson (2004) found clusters in genes including anti-apoptosis, cell proliferation and cell survival.

The main objective of this study is to test the hypothesis that LV integration alters the expression of RefSeq genes flanking the provirus site. The expression of these genes in liver tumours was compared to that of normal tissue and gender matched tumour samples. The second objective of this study was to investigate the possibility that multiple proviral insertions could induce HCC development.

4.3 Investigation of vector associated genotoxicity in mice following *in utero* delivery of EIAV, FIV and HIV vectors.

4.3.1 RT PCR analysis of tumour gene expression linking pLIONhAATGFP FIV derived vector insertions with alterations in gene expression

After retrieving FIV integration sites from the mouse tumours a study of the relationship between vector insertion and gene expression was investigated using RT PCR. RT PCR reactions were performed in normal and tumour DNA samples derived from FIV infected liver and gonad tissue. Preliminary amplification reactions of the GAPDH house keeping gene was used to measure the quality and integrity of the cDNA samples. RT amplifications showed clear bands that demonstrated the cDNA to be of high quality (Figure 19).

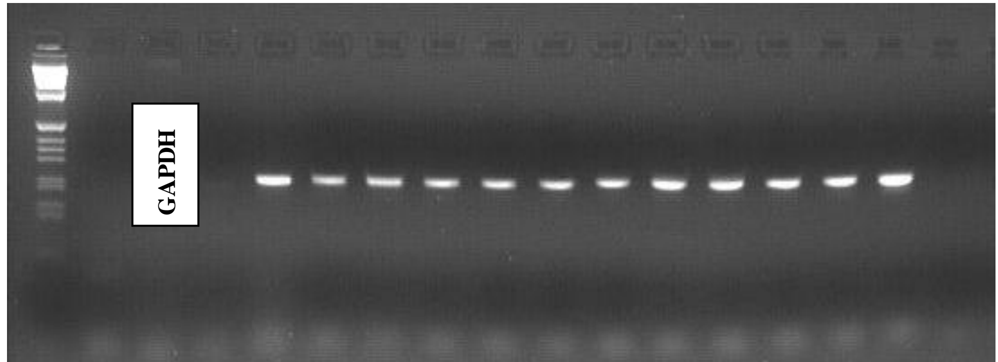
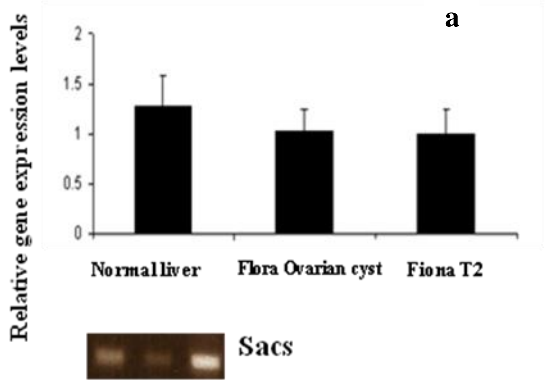
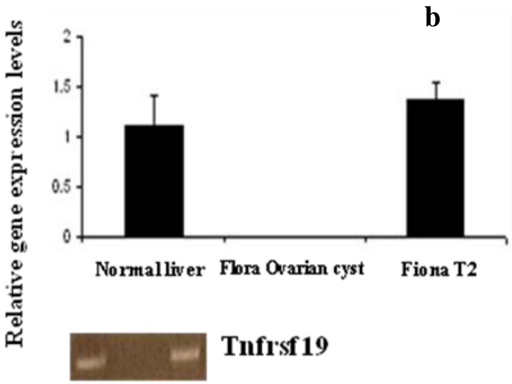


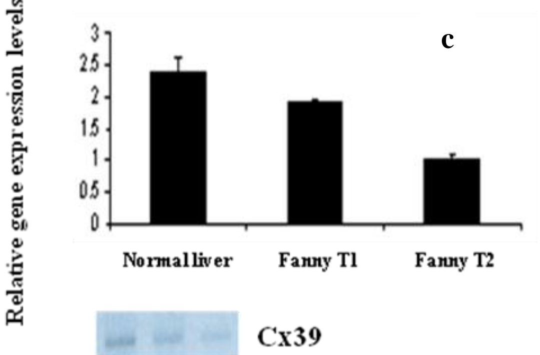
Figure 19. An example of GAPDH expression is shown for female normal liver (5, 9 and 13), Flora Ovarian Cyst (lane 6, 10 and 14), Fiona T1 (lanes 7, 11 and 15) and Fiona T2 (lanes 8, 12 and 16) samples.

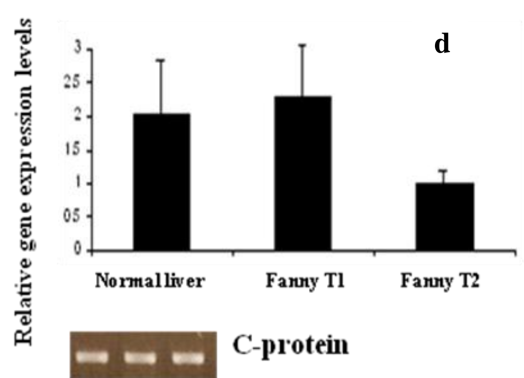
Band intensities between tumour samples and respective normal livers were used to reveal differences in gene expression levels. Relative gene expression levels were calculated by measuring the fluorescence peaks produced by single bands on an agarose gel. Each peak provided an area value. The average area value for each gene was taken and average area values were normalised against *GAPDH* area values to produce relative gene expression levels. In Flora ovarian cyst (20b) *Tnfrsf19* was found knocked down completely. Interestingly *Cx39* (Fanny T1), *Nek9* (Frank T1) also showed reduced expression in tumour samples containing insertions when compared to normal livers.

C-protein (Similar to Cyclin fold protein 1- humans), *Pscd3* and *Coro7* expression was increased when compared to gender matched normal samples. However, *Coro7* expression was also increased in Frank T2 samples. *Sacs* gene expression remained the same in samples with insertions and the tumour reference sample. However, *Sacs* expression in the tumour with insertion in this gene was lower than that in normal liver (Figure 20a).


a
Relative gene expression levels
2
1.5
1
0.5
0
Normal liver
Flora Ovarian cyst
Fiona T2
Sacs


b
Relative gene expression levels
2
1.5
1
0.5
0
Normal liver
Flora Ovarian cyst
Fiona T2
Tnfrsf19


c
Relative gene expression levels
3
2.5
2
1.5
1
0.5
0
Normal liver
Fanny T1
Fanny T2
Cx39


d
Relative gene expression levels
3
2.5
2
1.5
1
0.5
0
Normal liver
Fanny T1
Fanny T2
C-protein

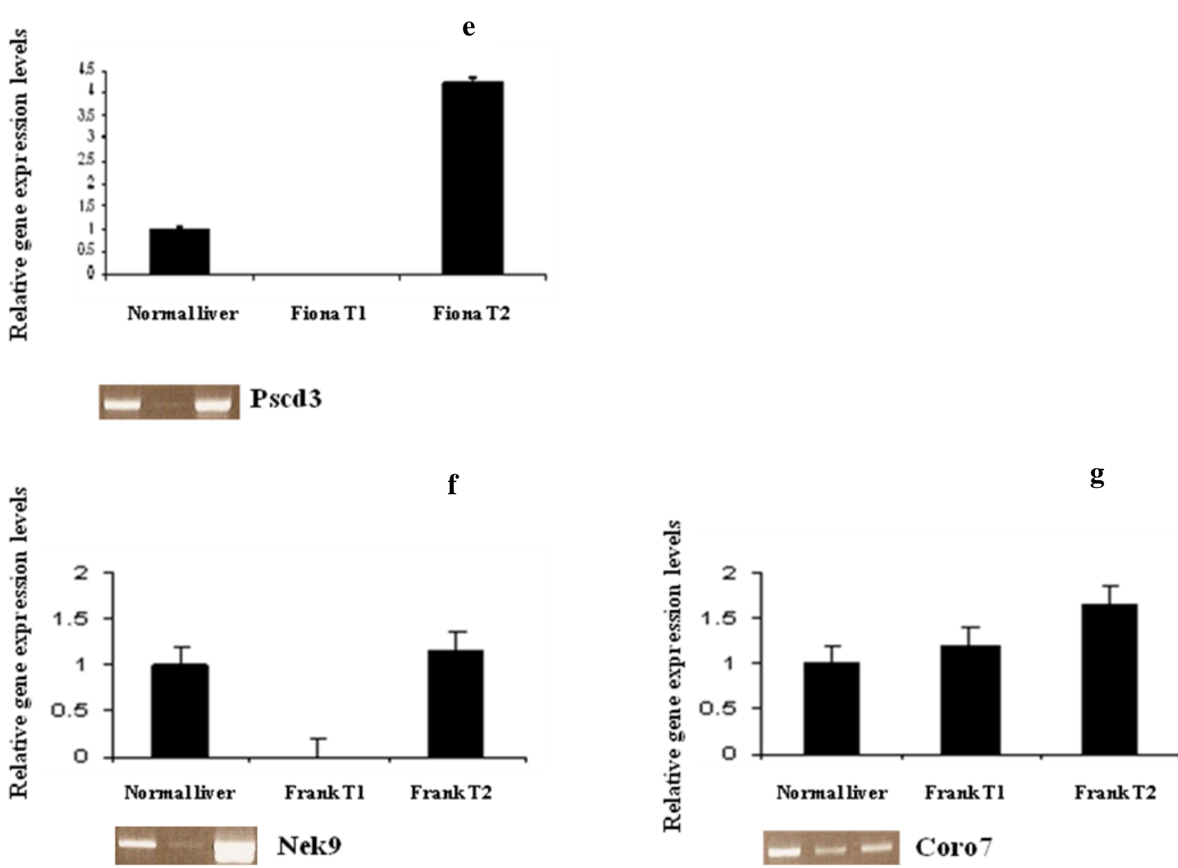


Figure 20 a – g. The expression levels of genes were measured in samples where FIV insertions were found. All gene expression amplification reactions were performed in triplicates. Average triplicate readings obtained from each insertion gene was normalised against the average reading obtained from *GAPDH* expression. Error bars represent standard deviation values obtained from triplicate readings. Gender matched normal liver and tumour samples were used as reference samples for changes in gene expression.

The association between FIV insertions and differential gene expression suggested FIV integrations could be involved in oncogenesis. QPCR is a highly sensitive assay also used to investigate differential gene expression and is an ideal tool to use to measure small changes in gene expression that RT PCR may not have identified. TaqMan probes and primer sets were obtained and used on the FIV infected samples. Controls consisting of gender matched normal and tumour samples were used for comparison of gene expression. In Frank T1 samples *Nek9* was found reduced when compared to normal liver (Table 45; P <0.05) and a gender matched tumour control sample (P <0.01). This reduction in expression was in agreement with that found for *Nek9* in Frank T1 by RT PCR.

QPCR also showed that *Pscd3* was up regulated in Frank T2 samples by 29%. This also agreed with RT PCR measurements. *Coro7*, *Nek9* and *Tnfrsf19* were downregulated in Tumour 1 samples by over 19%, 3 and 2% respectively and was also found by RT PCR. QPCR failed to identify the expression of *cx39* and *pdcd1lg2* in tumour samples and yielded very low (CT > 36) in normal liver samples.

RT PCR did not identify expression of *Pdcd1lg2* this suggests that this gene is expressed at very low amounts in this sample.

4.3.2 Real time PCR measurement of transcription levels of genes with FIV provirus insertions using QPCR analysis

Gene name	Tumour 1 +/-SEM	Tumour 2	Insertion change
FIV			
	Frank T1	**Frank T2**	
Nek9	0.11 (0.22)	3.78 (0.11)	Down-regulated
	Frank T2	**Frank T1**	
Coro7	0.11 (0.19)	19.48 (0.02)	Down-regulated
	Flora ovarian cyst	**Fiona T1**	
Tnfrsf19	0 (0)	2.69 (0.20)	Down-regulated
	Fiona T2	**Fiona T1**	
Pscd3	33.67 (0.06)	3.88 (0.02)	Up-regulated
	Fanny T1	**Fanny T2**	
Cx39	0 (0)	0 (0)	No change (no expression)
Pdcd1lg2	0 (0)	0 (0)	No change
Total deregulated	4		

Table 45. All gene expression experiments were performed in quadruplets and repeated on separate occasions. The conditions for all amplification reactions were optimised and a validation efficiency test was performed. *18sRNA* expression was used to normalise gene expression for all samples FIV infected samples. Tumour 1 and tumour 2 represent samples with and without insertions, respectively. Statistical analysis including the Student T test and 95% confidence interval testing was performed on all QPCR data. P. values were calculated for comparison between normal vs Tumour 1 and Tumour 1 vs Tumour 2.

4.3.3 QPCR analysis in EIAV infected samples

EIAV insertion genes were also observed for changes in gene expression to determine whether this pattern also occurs for genes carrying EIAV insertions. Several insertion genes showed changes in their tumour of origin (described in Table 45 and 46 as tumour 1). *Park7* was identified as a hotspot for EIAV insertion genes in previous (Themis *et al*, 2005) and current data described in this study. This putative oncogene is known to be over expressed in cancer (Nagakubo *et al*, 1997; Sitaram *et al*, 2009; Vasseur *et al*, 2009). Upregulation of *Park7* was seen in the tumour without insertion but in tumour 1 was downregulated when measured by QPCR and in the microarray of this tumour (Table 46). Preparation of RNA and collation of raw data for microarray analysis was conducted by Dr Manfred Schmidt (National Center for Tumor Diseases and German Cancer Research Center, Heidelberg, Germany). *Acvr2a* belongs to a family of activins involved in transforming factor beta (TGFβ) signaling. Disregulation of activins such as *Acvr2a* has been shown in diseases such as cancer (Chen *et al*, 2006; Li *et al*, 2006; Tsuchida *et al*, 2009). This gene was upregulated by 13% in tumour 1. This level of gene expression was higher than that found in normal liver ($P<0.05$) but lower than the tumour without the insertion ($P <0.01$). The *Pah* gene was shown to produce a high sequence count in Adame T1 samples. QPCR showed down regulation of this gene in the tumour containing the insertion when compared to normal liver ($P <0.05$) and the tumour without the insertion ($P <0.01$). Reduced expression of the *Pah* gene was also found by microarray of this tumour (Table 46).

Rabgef1 and *Mrlp23* were both upregulated in tumour 1 (Jonas Large tumour) and tumour 2 (Table 49). Upregulation of this gene was also observed in microarray analysis (Table 49).

The expression of *Park7*, *Rabgef1*, *Mrlp23* and *Uvrag* was measured in additional male mouse uninfected liver samples (Cat and Uninfected liver sample shown in Table 49; Appendix 4) and a gender matched FIV infected tumour (Frank T2). These additional reference samples (shown in red) produced variable changes in gene expression when compared to tumour 1 (Jonas large tumour).

Bre plays a role in the DNA damage response and repair pathways and in inflammation. QPCR showed this gene to be down regulated in the tumour 1 sample (Jonas large tumour) when compared to normal (P <0.05) and the tumour without insertion in this gene (P<0.01). Deregulation of *Bre* was also shown in microarray of Jonas large tumour samples (Table 46). *Mark3* and *Katna* were not analysed by microarray analysis however QPCR showed upregulation and downregulation of each gene, respectively, in tumour 1 when compared to normal liver (P <0.05) and the tumour without insertion in this gene (<0.01).

Mrpl1 and *Setd5* showed high sequence counts of 743 and 267, respectively. Microarray data (not shown) revealed upregulation of *Mrpl1* but showed insignificant changes (P >0.05) in gene expression for *Setd5*. QPCR analysis was also performed on these genes, however, the integrity of RNA in Gerry tumour 1 samples appeared reduced due to the occurrence of degradation as found by gel electrophoresis (data not shown). The poor integrity of this sample was also reflected by high *18sRNA* CT values (> 16 cycles). Consequently QPCR could not be performed reliably on these genes.

Gene name	Tumour 1 +/-SEM	Tumour 2 +/-SEM	Insertion change*	Microarray
EIAV				
	Adame T1	**Fiona T2**		
Acvr2a	1.30 (0.19)	10.79 (0.07)	Up-regulated	Nr
Pah	0.35 (0.18)	3.70 (0.06)	Down-regulated	-0.19
	Adame T1	**Fanny T2**		
Hnf1a•	0.55 (0.10)	0.20 (0.26)	Down-regulated	Nr
Hnf4a•	0.62 (0.10)	0.18 (0.28)	Up-regulated	-0.53
Foxa2•	1.57 (0.06)	0.75 (0.23)	Up-regulated	Ns
	Jonas large tumour	**Jonas small tumour**		
Park7	0.55 (0.04)	502.61 (0.24)	Down-regulated	0.39
	Cat	**Uninfected liver**		
	0.31 (0.17) ↓	3.04 (0.12) ↑		
	Frank T2			
	2.71 (0.07) ↑			
	Jonas large tumour	**Jonas small tumour**		
Rabgef1	2.32 (0.05)	27.87 (0.13)	Up-regulated	Ns

	Cat	Uninfected liver		
	0.23 (0.10) ↓	4.69 (0.16) ↑		
	Frank T2			
	1.80 (0.41) ↑			
	Jonas large tumour	**Jonas small tumour**		
Mrpl23	5.11 (0.22)	27.53 (0.08)	Up-regulated	0.74
	Cat	Uninfected		
	0.40 (0.15) ↓	6.12 (0.06) ↑		
	Frank T2			
	14.93 (0.28) ↑			
Uvrag	**Jonas large tumour**	**Jonas small tumour**		
	1.28 (0.20)	17.85 (0.13)	Up-regulated	0.78
	Cat	Uninfected liver		
	0.21 (0.18) ↓	2.64 (0.12) ↑		
	Frank T2			
	2.84 (0.05) ↑			
	Jonas large tumour	**Monika tumour**		
Bre	0.50 (0.04)	0.99 (0.07)	Down-regulated	-0.12

	Gina T1	Monika tumour		
Mark3	1.44 (0.53)	1.13 (0.03)	Up-regulated	Nr
Katna	0.81 (0.06)	0.57 (0.11)	Down-regulated	Nr
Total deregulated 12				

Table 46. Gene expression ($2^{\Delta\Delta CT}$) were calculated for each gene and compared to gender matched tumour and normal samples. All of the top differentially expressed genes (n = 500). These differentially expressed were subjected to Benjamini Hochberg correction to allow multiple comparisons of microarray data. Only genes that showed significant (P <0.05) after Benjamini Hochberg correction were selected. Nr and Ns were used to describe insertion genes where microarray data was not available and genes where p. values obtained were above significant levels (<0.05) respectively.

4 genes (*Pah*, *Park7*, *Bre* and *Katna*) showed reduced levels of expression following EIAV vector integration into these genes. *Acvr2a* appears upregulated when compared to normal liver samples. However, results show that *Acvr2a* was also downregulated in tumour 1 when compared against a tumour control. Several genes (*Rabgef1*, *Mrpl23*, *Uvrag* and *Mark3*) were also up regulated in tumour 1.

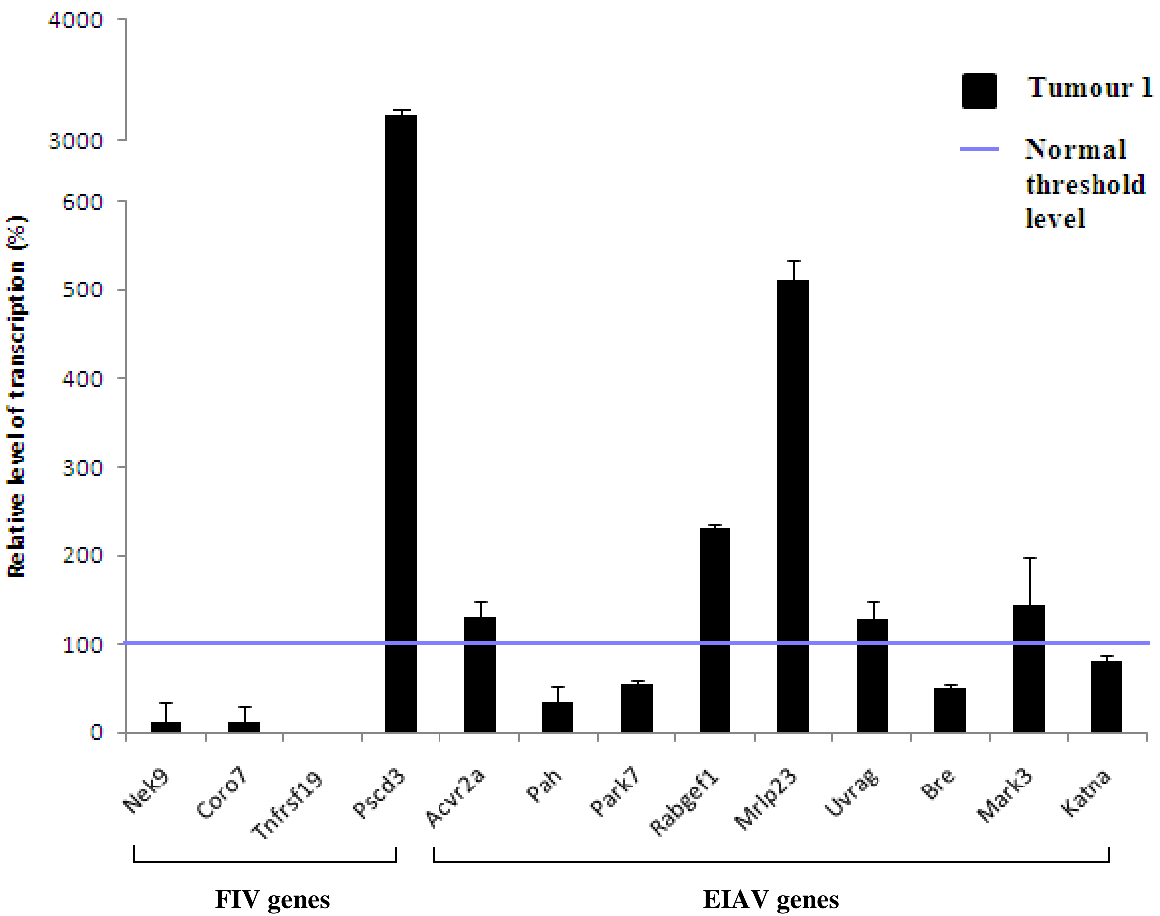


Figure 21. Histograms were used to represent gene expression data ($2^{-\Delta\Delta CT}$). Error bars represent standard error of the means taken from quadruplet readings. Relative gene expression data for all normal liver samples were taken as 100%. The relative quantification (RQ) levels calculated for calibrator samples (normal liver sample) are equal to 1. The blue line is used to represent 100% RQ levels in calibrator samples.

4.3.4 Analysis of FIV insertion gene pathways by 'STRING'

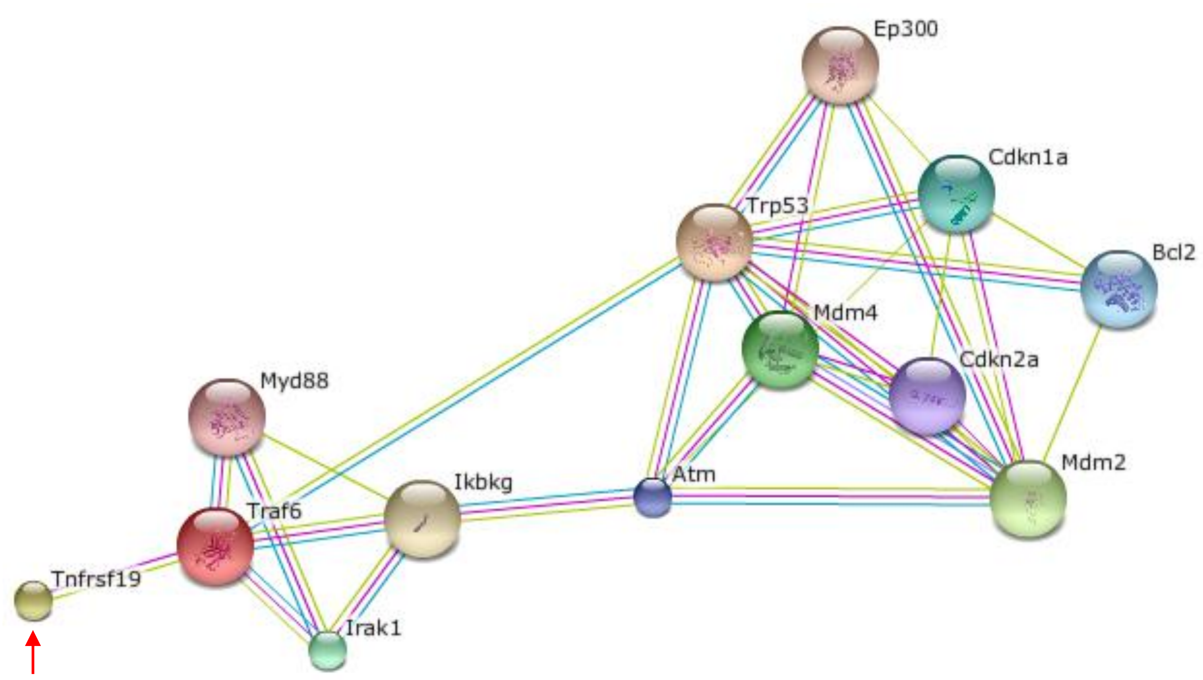
FIV insertions that showed differential expression by QPCR and microarray analysis were investigated to find networks of interacting genes in pathways that are known to be implicated in oncogenesis and HCC development. To do this, inserted genes were investigated using STRING (string-db.org/) analysis. STRING provides networks between proteins obtained from literature as images. For example *Tnfrsf19* and *Pscd3* were found to associate with the transformation related protein 53 - *Trp53* (Figure 22 a-b).

Trp53 is a known tumour suppressor gene that regulates genes involved in several mechanisms including DNA repair, apoptosis and cell cycle control. *Tnfrsf19* is a functional partner closely related to the TNF receptor associated factor 6 (*Traf6*) and is an important member of the TNFR superfamily. This gene also functions as a signal transducer of NF-kappa B genes involved in inflammation.

Traf6 interacts with several cancer associated genes involved in cell cycle progression in response to DNA damage and cellular proliferation (*Cdkn1a, Cdkn2a and Bcl2*) and the *Atm* gene that is involved in DNA damage and repair mechanisms. *Tnfrsf19* expression was found to be reduced in the Flora ovarian cyst tumour sample suggesting this gene to be possibly involved in oncogenesis.

Pscd3 (*Cyth3*) associates closely with the RAS superfamily member ADP-ribosylation factor 6 (*Arf6)* (Figure 22b). *Arf6* interacts with the protoocogene *v-akt* murine thymoma viral oncogene homolog 1 (*Akt1*). Moreover *Akt1* relates to the key tumour suppressor gene phosphatase and tensin homolog *(Pten*). This suggests that *Pscd3* is an important member of an oncogenesis pathway that may lead to HCC. *Nek9* and *Coro7* were also investigated in STRING, however no pathways were found relating these genes to other cancer associated genes.

a



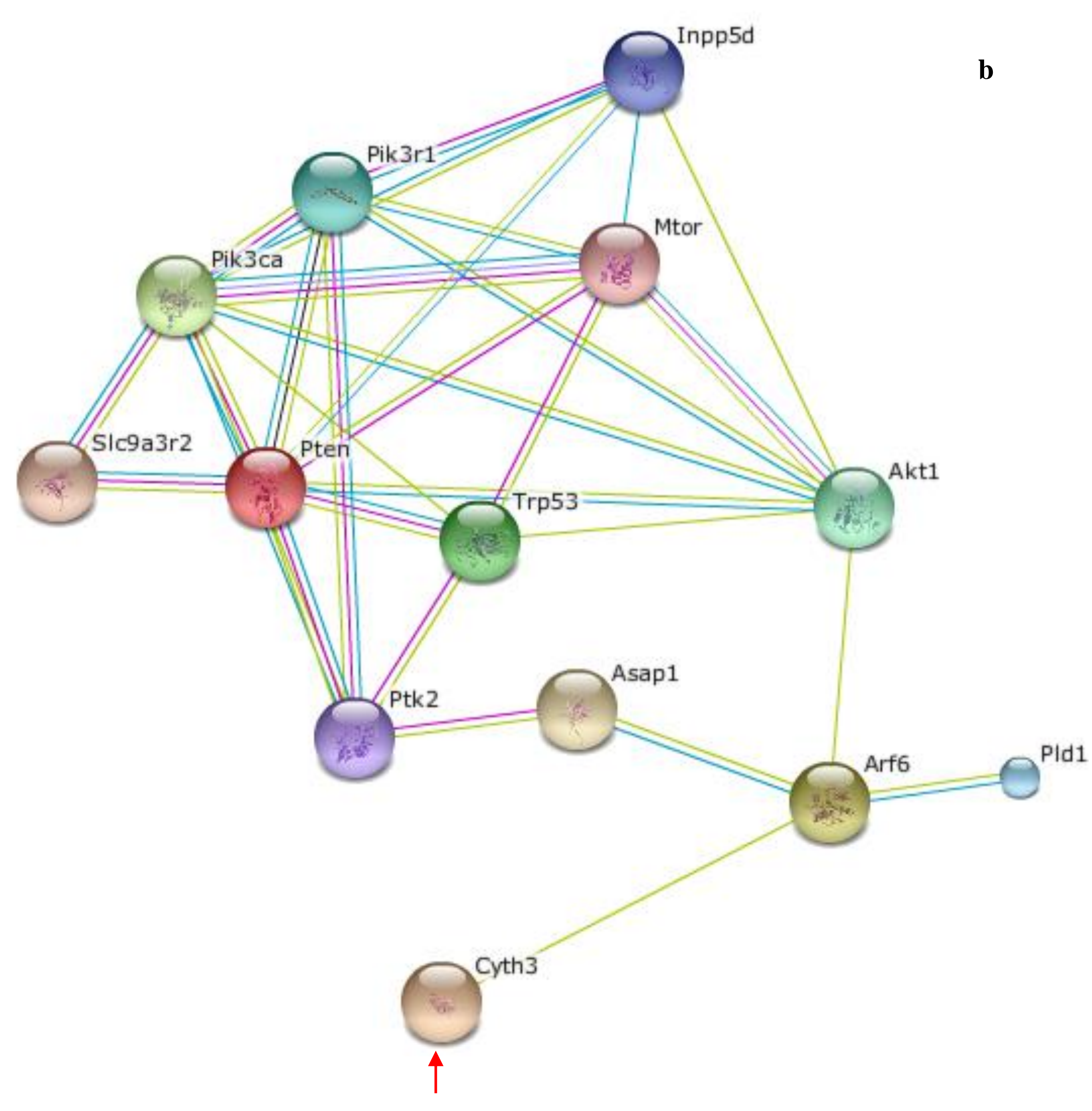
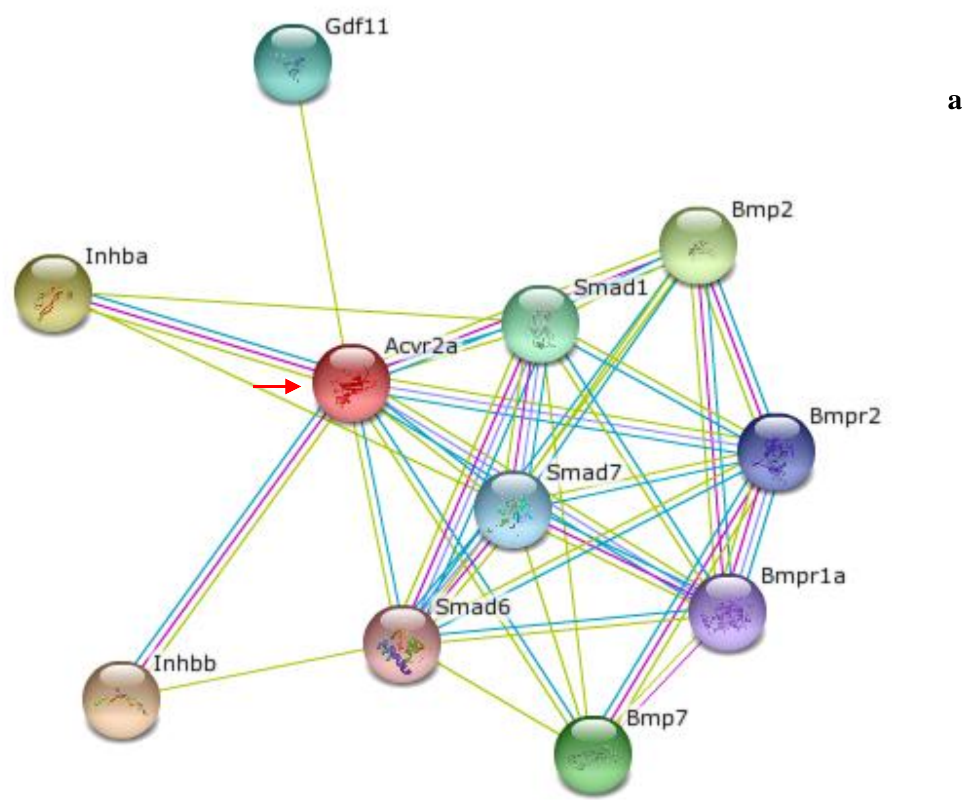


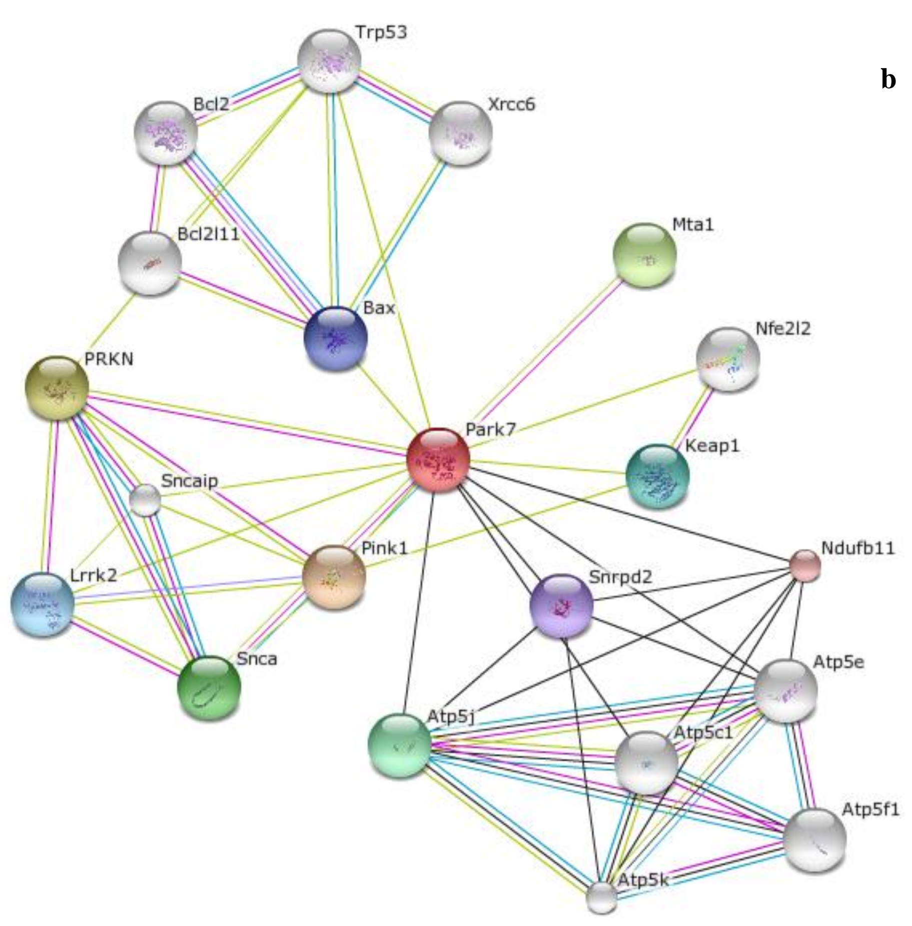
Figure 22 a-b. STRING analysis was used to predict protein-protein interactions between FIV insertion genes. The STRING database was also used to identify cancer associated genes. QPCR analysis was used to investigate insertion genes that interacted with known cancer associated genes or formed part of an oncogeneic network.

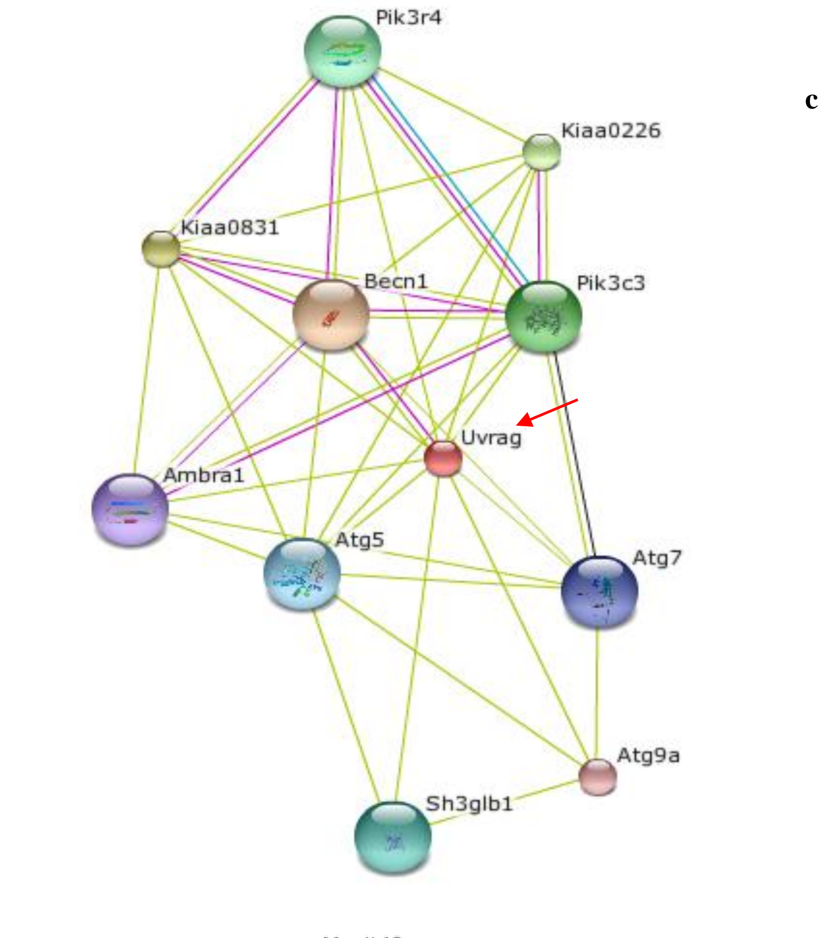
4.3.5 Analysis of EIAV insertion gene pathways using the STRING database and microarray analysis.

EIAV insertion genes that were represented as clonal in tumours by high 454 sequence counts with altered expression by QPCR and microarray analysis were entered into the STRING program as was performed for FIV inserted genes. *Park7*, *Uvrag*, *Mrlp23*, *Acvr2a, Bre* and *Pah* were each found associated with cancer genes. *Acvr2a* was found to share common activation pathways with *Smad* proteins including *Smad1, 6 and 7*. (Figure 23a) *Inhiba/b* genes are also known to be functional partners of *Acvr2a. Smad 7* expression was not investigated in Adame T1 samples. However, the P. Value for *Smad1, Inbba and Inhibb* was above the significance value (<0.05) used to correct samples. *Park7* was found to be associated with the apoptotic regulator gene *Bax* (BCL2- associated X protein) and the tumour suppressor gene Trp53 (Figure 23b). *Bax* is a functional partner of *Trp53* and the oncogene *Bcl2*. Interestingly, *Bax* was found to be upregulated by 18.8% in Jonas Large tumour compared to the normal liver control. The tumour suppressor gene *Becn1* was also found to be associated with *Bax*. The tumour suppressor gene *Becn1* was found to be associated with *Uvrag* (Figure 23d). Microarray also showed *Becn1* was upregulated by 2.1% in Jonas large tumour however, another important gene included in this network is *Pik3c3*. *Pik3c3* plays a key role in protein sorting and the mTOR signaling pathway and was downregulated by 17.6% by microarray. Cancer associated genes found in the *Bre* protein-protein network also include *Smad* and *Brca1*. Interestingly the proto-oncogene - *Hras* (v-Ha-ras Harvey rat sarcoma viral oncogene homolog) was also found in the STRING network of *Mrpl23*. This gene was not found in the microarray but shows the potential link between EIAV integration genes and oncogenesis.

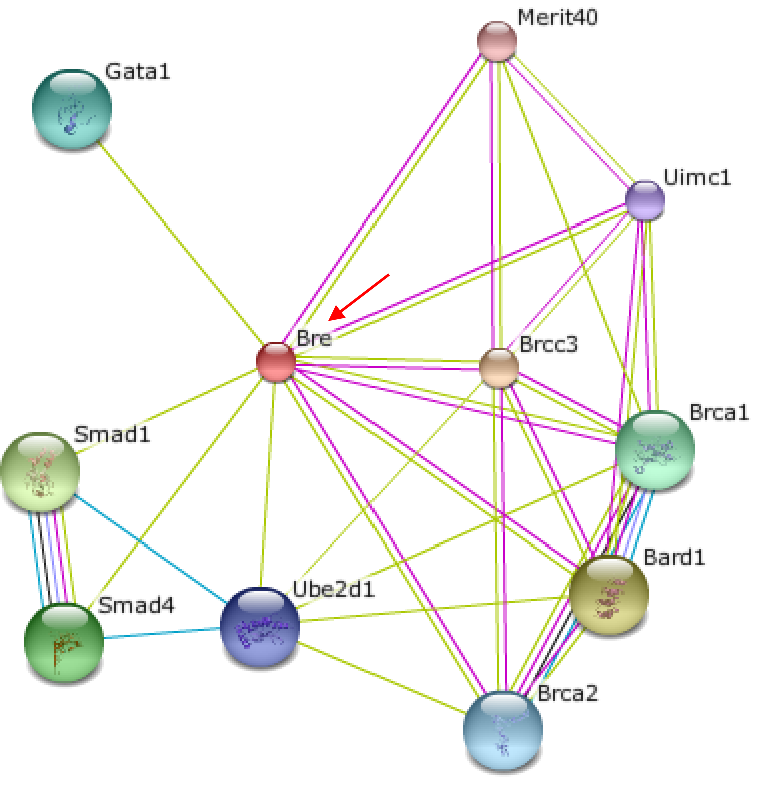
The *Pah* gene was shown to have a close interaction with hepatic nuclear factor 1 alpha (*Hnf1a), Hnf* 4 alpha (*Hnf4a*) and fork head box a2 (*Foxa2*) (Figure 23f). Each gene was shown to be altered with QPCR analysis. Because *Hnf4a* is reduced in HCC and liver cancer, QPCR was performed on this gene and this revealed a reduction in its expression by 62% (P. Value <0.05) and a reduction of *Hnf1a* by 55 % (P. Value <0.05) in Adame T1 tumour sample. Interestingly, a 15.7% increase in *Foxa2* expression was found in Adame T1.


a
Gdf11
Inhba
Acvr2a
Smad1
Bmp2
Bmpr2
Smad7
Bmpr1a
Smad6
Inhbb
Bmp7

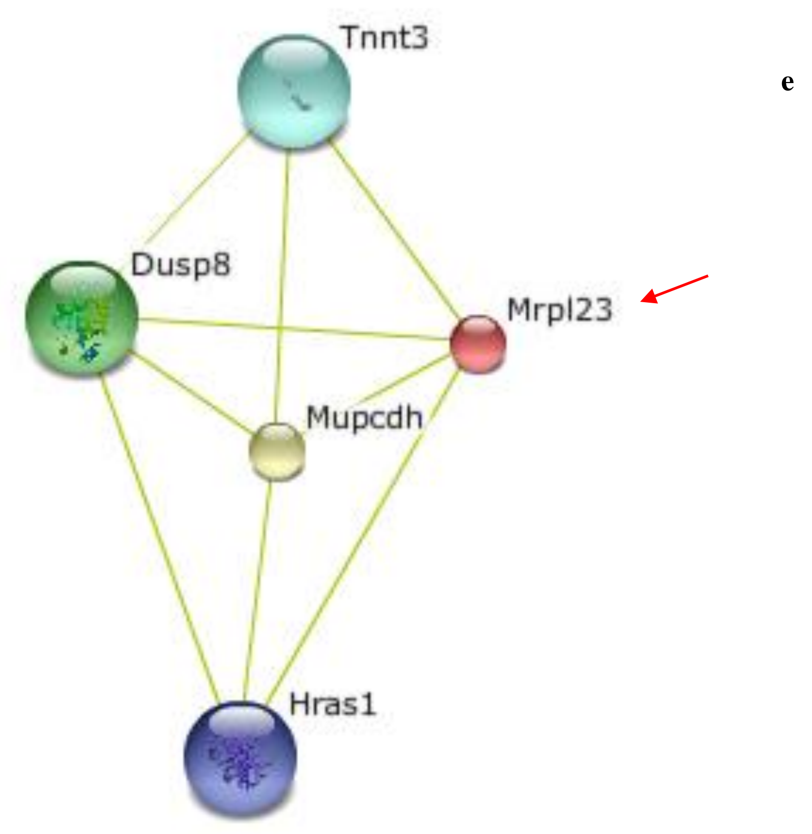

b
Trp53
Bcl2
Xrcc6
Bcl2l11
Mta1
Bax
Nfe2l2
PRKN
Park7
Keap1
Sncaip
Pink1
Ndufb11
Lrrk2
Snrpd2
Snca
Atp5e
Atp5j
Atp5c1
Atp5f1
Atp5k



c



d


Tnnt3
e
Dusp8
Mrpl23
Mupcdh
Hras1

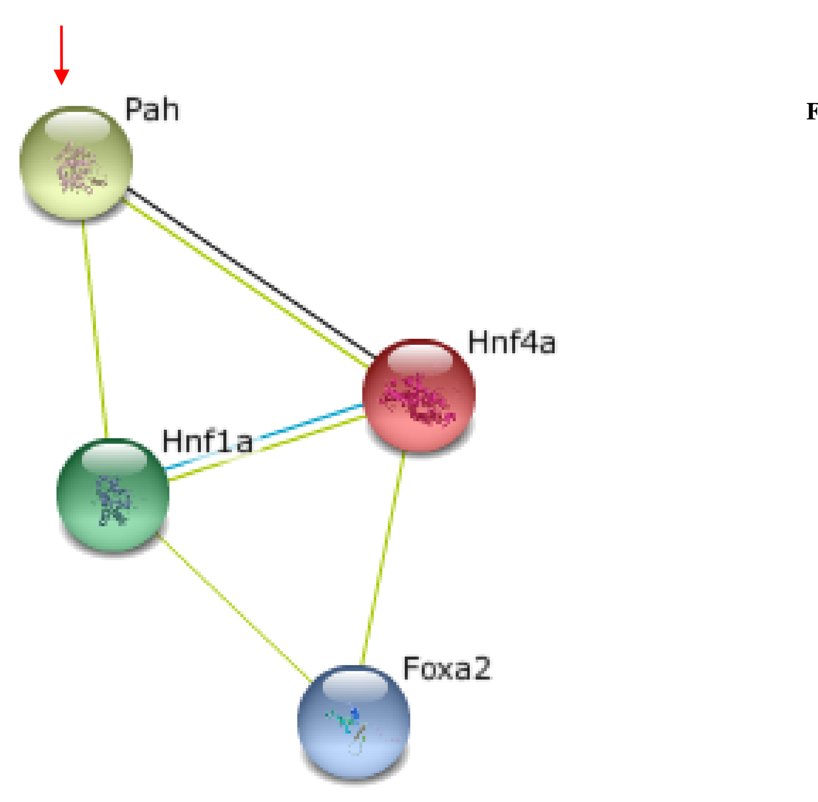

Pah
F
Hnf4a
Hnf1a
Foxa2

Figure 23 a-f. Gene pathways associated with genes flanking the EIAV provirus. Several SMART2 insertions were found in this study. 454 sequencing revealed 7 unique hot spots (repeated insertions in the same tumour) for *Acvr2a*, *Park7*, *Uvrag* and *Bre*. These genes were found to be differentially expressed in the tumour of insertion using QPCR and microarray analysis. Protein-protein interaction between SMART2 insertions was investigated using the STRING database. STRING analysis revealed important interactions with cancer associated genes. A) *Acvr2a*, B) *Park7*, C) *Uvrag*, D) *Bre* E).

IPA analysis software was used to demonstrate the role of insertional mutagenesis and its role in HCC development in EIAV infected mice. Q-PCR and microarray showed that *Uvrag* and *Park7* had differential gene expression in Jonas large tumour. STRING analysis also showed that the interaction between *Bax* and *Trp53* is an important link connecting *Uvrag* and *Park7* with several cancer associated genes (Figure 24). IPA analysis showed that *Uvrag* form part of a oncogenic network with important cancer genes including *Trp53*, *Atg7* and *Bcl2* which are involved in proliferation of hepatocytes, tumorigenesis of mice and liver cancer respectively (Figure 25). This supports the role of EIAV integration into *Park7* and *Uvrag* as possibly contributing to HCC. Fisher's exact test was conducted to determine the statistical significance of genes enriched within a given GO category and validate genes overrepresented in data sets. P-values$<10^{-20}$ or lower were used to select highly significant overrepresentation after Benjamini Hotchberg correction.

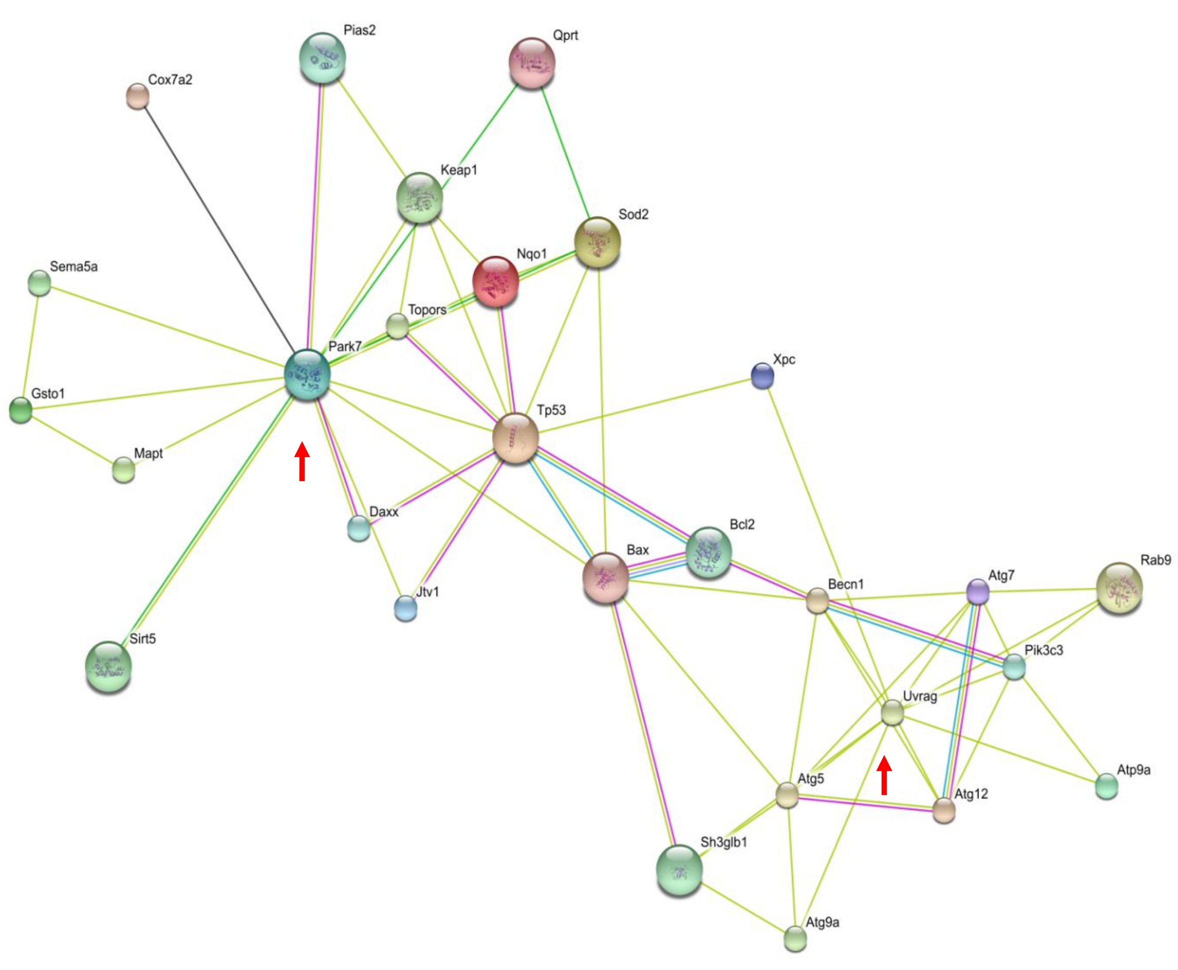


Figure 24. The relationship between *Uvrag* and *Park7* demonstrated using STRING analysis. *Park7* interacts with *Uvrag* through a common network of genes. This includes important cancer associated genes such as *Tp53*, *Bax* and *Bcl2*.

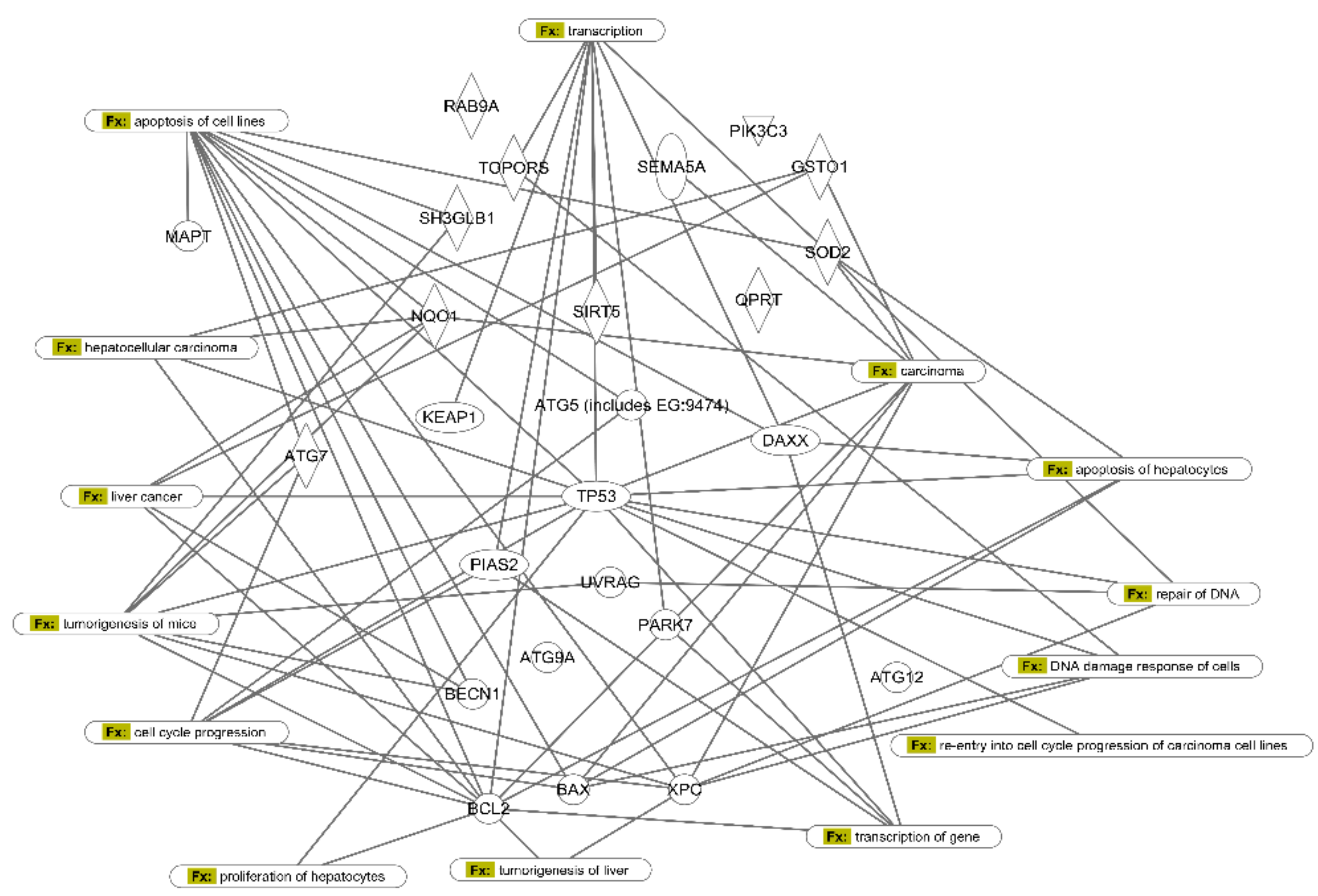


Figure 25. IPA analysis: Cooperation between *Park7* and *Uvrag* was found using IPA analysis. Both genes were identified in the Jonas large tumour. Black lines represent interaction between *Park7*, *Uvrag* and their associated genes. Boxes are used to represent key biological processes and mechanisms associated with this pathway. IPA analysis demonstrates the importance of these genes in liver cell growth, cell cycle progression and liver cancer and HCC.

4.3.6 The effects of SMART2Z integration on normal splicing of the *Pah* gene

In Adame T1 tissue the SMART2Z vector was found in intron 2 of the *Pah* gene (Figure 26). As aberrant splicing has been shown to occur between integrated vector sequences and host genes (Coffin, Hughes and Varmus, 1997; Nienhuis *et al*, 2006) this potential was investigated for SMART2Z and the *Pah* gene as a possible cause for the reduction in *Pah* gene expression. An *in silico* construct was designed in collaboration with Dr Evgeny Makarov (Brunel University, Uxbridge, UK) to represent the virus integration site (see Appendix 11). Amplification reactions were designed to identify the junction between the upstream region of intron 2 and the SMART2Z vector sequence (Figure 26).

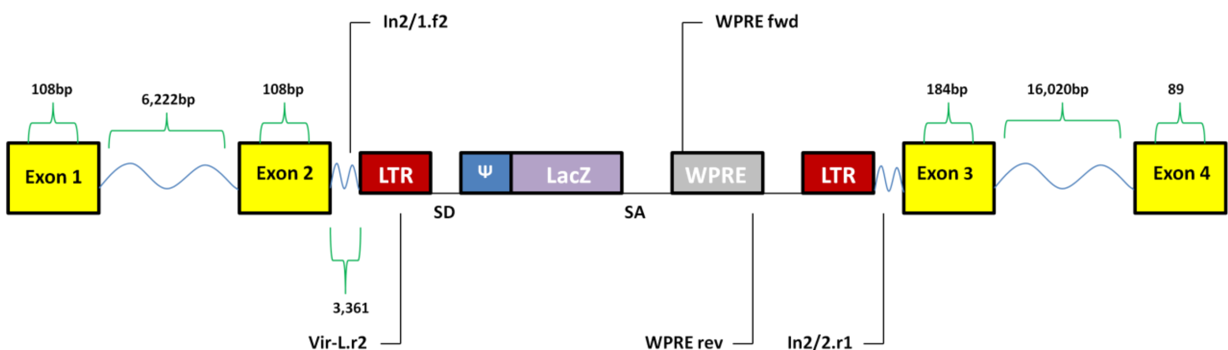


Figure 26. Predictions of the posssible splice combinations were made to repersent normal and abnormal splicing between the vector and all 13 exons found in the *pah* gene. (see Appendix 11). Predicted fragments were generated taking into account the presence of splice donor (SD) and splice acceptor (SA) sequences present in the SMART2Z vector construct. Oligonucleotide primers were designed to amplify genomic and cDNA samples including vir-L-R2, In2/1.f2, Wpre forward and reverse sequences and In2/2.R1.

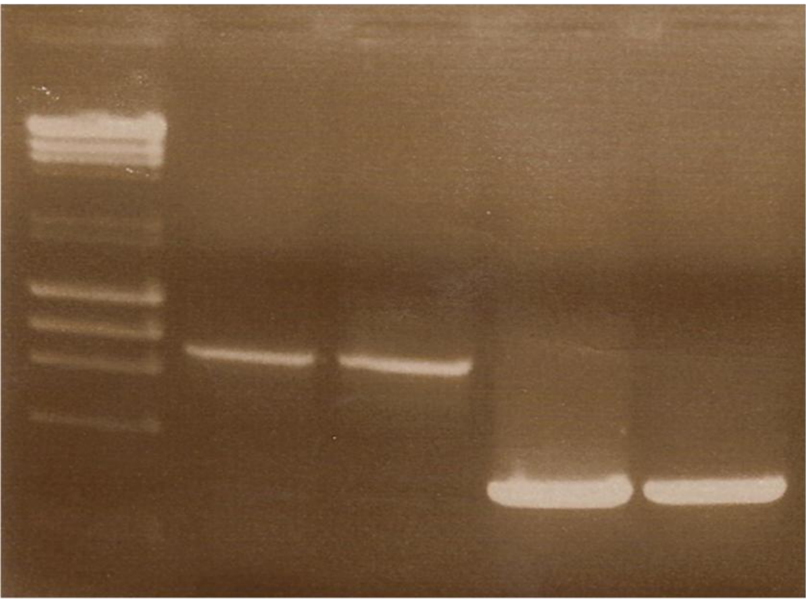


Figure 27. Several PCR reactions were designed to identify aberrant splicing leading to virus/host mRNAs between SMART2Z and the *Pah* gene. RT PCR was used to amplify the junction (1010bp) between the downstream region of intron 2 and the SMART2Z sequence using in2/2.r1 and WPRE foward oligo sequences (lanes 1 and 2). A 485bp region was amplified using in2/r.f2 and vir-l.r2 oligo sequences (lanes 3 and 4). Clear bands demonstrated the presence of these regions. Ampliflication of the WPRE domain was performed as a control to confirm the presence of vector in infected samples. A 356bp reaction was shown on a gel (data not shown). All amplification reactions were sequenced by Dundee sequencing service.

RT PCR amplification reactions were performed using oligo sequences that hybridise to the *Pah* gene (Figure 27). PCR amplification products were purified and sequenced (see Appendix 11). These results confirmed the presence of exons 1-13. However, these results failed to identify viral sequences attached to the *Pah* gene cDNA in Adame T1. This suggests that following viral integration normal splicing was not altered Adame T1. However, this does not rule out that splicing between *Pah* and SMART2Z was not detected.

5.1 The relationship between viral integration and E2F target genes

Eukaryotic and viral gene expression may be regulated by transcription factors, regulatory elements, promoters or epigenetic modifications (Bartholomew and Ihle, 1991; Nevins 1992; Weber and Cannon, 2007; Roy, Sen and Roeder, 2011). Interaction between E2F transcription factors and E2F responsive promoters is one mechanism used to regulate gene expression (Chen, Tsai and Leone, 2009; Singh *et al*, 2010).

The E2F superfamily consists of eight proteins (E2F 1-8); the structure of all E2F proteins (excluding E2F7 and 8) consists of an E2F component and a DP1/2 subunit (Fig 28; Bracken *et al*, 2004; Carcagno *et al*, 2011). E2F transcription factors bind to specific DNA sequences (‘TTTCGCGC’) located in the promoter region of E2f target genes (Nevins, 1992; Shirodkar *et al*, 1992; Tao *et al*, 1997). The adenovirus was originally used to investigate the activity of E2F transcription factors over 20 years ago (Kovesedi *et al*, 1986; Yee *et al*, 1987; Levine *et al*, 2009; Chen, Tsai and Leone, 2009). These studies reported specific binding between the adenoviral gene promoter and cellular factors in adenovirus infected cells.

E2f target genes are involved in several biological processes including cell cycle control (Cam and Dynlacht, 2003; Blais and Dynlacht, 2004; Chen *et al*, 2009; Carcagno *et al*, 2011), DNA damage and repair mechanisms (Field *et al*, 1996; Huang *et al*, 1997; Stiewe and Pützer; 2001; Hsieh *et al*, 2002) and cellular differentiation (Nevins, 1992; Field *et al*, 1996; DeGregori *et al*, 1997; Riz and Hawley, 2005). E2F family members also demonstrate dual functions including the ability to exert both oncogenic and tumour suppressor activity (Table 47; Bracken *et al*, 2004). This dual activity allows E2F to take part in several mechanisms involved in cancer development (Müller and Helin, 2000; Nevins, 2001).

5.1.1 E2f and *Rb1* interaction is fundamental to E2F activity

Early observations reported E2F as the functional link between the tumour suppressor gene - *Rb* (retinoblastoma) and the E1A (early region 1 A) oncoprotein present in the adenovirus genome (Chellapan *et al*, 1992; Nevin, 1992). In 1991 Chellapan *et al* reported that increased adenovirus infection correlated with the dissociation of the E2F-Rb complex (Chellapan *et al*, 1991; 1992). *Rb* is a key tumour suppressor gene that regulates the G1 to the S phase of the cell cycle DeGregori *et al*, 1997; Stevens and La Thangue, 2003). The RB proteins (pRB) are collectively known as pocket proteins that include p107 and p130 (Lee *et al*, 2002). In G0-G1 cells pRB binds to E2F family members and suppress E2F activity (Lee *et al*, 2002). Mitogenic signals activate pRB and induce phosphorylation of pRB (Lee *et al*, 2002). As a result the E2F-RB complex dissociates, E2F is released and allowed to promote the transcription of E2F target genes (Stevens and La Thangue, 2003; Bracken *et al*, 2004; Calzone *et al*, 2008). Several studies have shown that *Rb* interacts with oncoproteins and is frequently mutated or deleted in cancers such as HCC (Ceccarelli *et al*, 1998; Lohmann *et al*, 1997; 1999; Müller *et al*, 2001; Nevins, 2001; Classon and Harlow, 2002; Calzone *et al*, 2008; Ozturk *et al*, 2009; Reed *et al*, 2009; Bremner and Zacksenhaus, 2010).

Subgroup	E2f members	Function
1	E2f1-3	Interacts with pRb Activates gene expression
2	E2f4-5	Interacts with pRb Represses gene expression - Rb dependant activity
3	E2f6-8	Represses gene expression - Rb independent activity

Table 47. E2f family members can be categorised into three major groups according to their sequence homology and biological activity.

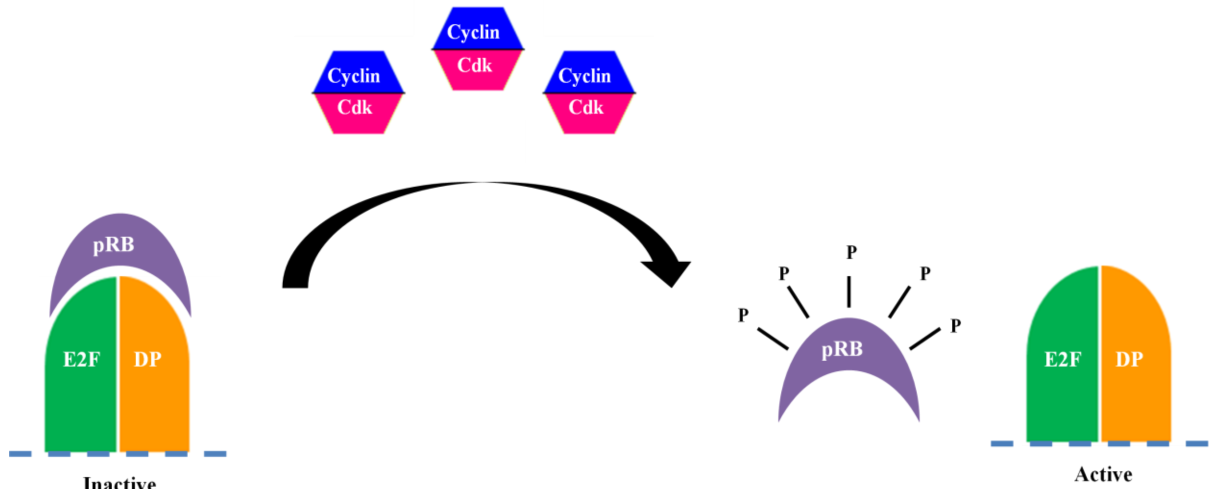


Figure 28. E2F mediated activation of cellular genes. Dissociation of E2F: pRB complexes allow E2F to perform transcriptional activation. DP (E2F dimerisation partner) represents a heterodimersation domain that binds to E2F family members. DP allows E2F to bind to DP1 and DP2 proteins and interacts with RB1 and RB related proteins. Cyclin/Cdk complexes interact to regulate cell cycle progression and the activity of pRB and E2F proteins. In normal cells pRb is hypophosphorylated. External signals act on cyclin and CDK complexes. These active complexes phosphorylate the *Rb* gene; this in turn cause the release of E2F.

5.1.2 E2F and DNA damage

During DNA damage responses E2F is able to regulate the cell cycle, apoptosis and associate with DNA repair proteins (Figure 29; DeGregori *et al*, 1997). DeGregori *et al*, (1997) inserted E2F family members into an adenoviral construct. These recombinant viral constructs were used to infect REF52 cells. Results showed that E2F family members were able to induce apoptosis and suggests that E2F1 may function as a signal for the induction of the apoptotic pathway. E2F1 has been shown to play an important role in controlling cell cycle control and apoptotic cell death in response to DNA damage and oncogene activation (Carcagano *et al*, 2009). Mutations in E2F1 results in the suppression of apoptosis (Tsai *et al*, 1998) and over expression of E2F effectively promotes apoptosis and subsequent malignant growth (Pierce *et al*, 1999). Carcagano *et al* (2009) investigated the effect of genotoxicity on E2F in a human fibroblast cell line. Results showed that E2F1 is over expressed in response to genotoxic stress in an ATM/ATR dependant manner. Whilst other studies report that E2F1 can suppress apoptosis (Denchi and Helin, 2005). Huang *et al* (1997) also reported the activation of E2F1 during DNA damage and apoptosis.

Other studies have also demonstrated a link between DNA damage and other E2F family members such as EF2 and E2F3 (Martinez *et al*, 2010).

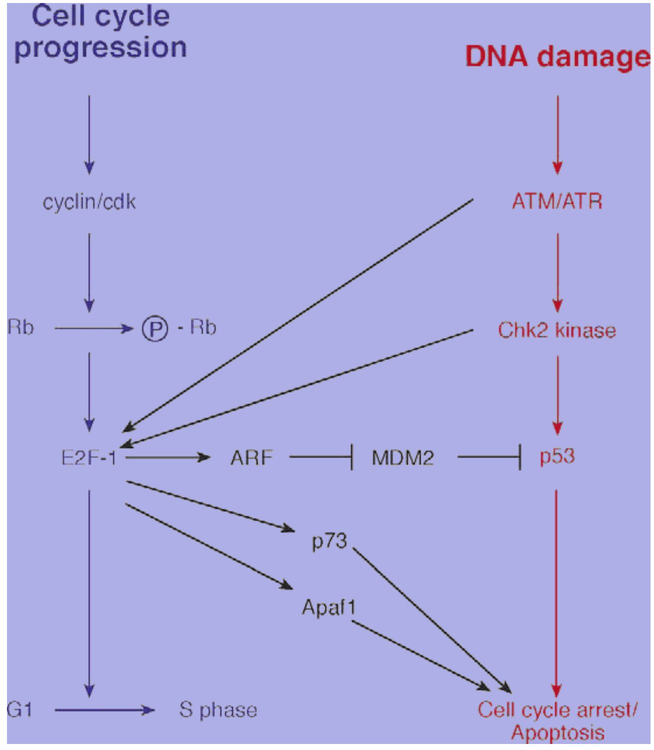


Figure 29. E2f plays a role in the DNA damage pathway. This image shows the close interaction between components involved in the cell cycle and DNA damage. Image taken from Stevens and La Thangue, 2003.

Aberrant regulation of E2F activity may facilitate tumour progression by activating the mutation of crucial tumour suppressor genes such as *Rb* and *p16ink4a*. Whilst E2F mutations exist these are not as frequent as the deregulation of E2F target genes in tumours. Therefore deregulation of E2F targets may provide a useful marker of cancer development. Coordination exists between E2F regulation and epigenetic mechanisms. E2F binding sites contain 1 or 2 CPG dinucleotides that are potential targets for regulation by DNA methylation (Campanero, Armstrong and Flemington, 2000). Campanero *et al* (2000) investigated the effect of CPG methylation occurring in the promoter region of several E2F target genes.

Velasco *et al*, 2010 demonstrated a functional relationship between *Dnmt3b* and the transcriptional repressor *E2F6* in *Dnmt3b* mutant mice. This study showed that

Dnmt3b target genes were dependant on the binding of *E2F6*. Moreover, E2F6 has been shown to interact with repressive chromatin-modifying enzymes which play a role in methylation & epigenetic regulation of transcription.

5.2 Epigenetic modification of host genes following lentiviral vector delivery: TSG methylation in HCC and non tumour liver

5.2.1 Epigenetic regulation of transcription

The term epigenetics is used to describe stable heritable changes in gene expression that have no affect on the nucleotide sequence of DNA (Jaenisch and Bird, 2003; Jiang *et al*, 2004; Rodenhiser and Mann, 2006). The biological significance of epigenetics in normal development, cell proliferation, gene expression and the aeitology of disease have become increasingly apparent since epigenetics was first introduced by Conrad Waddington in the 1940′s (Waddington, 1942; Devaskar and Raychaudhuri, 2007; Matouk and Marsden, 2008).

Epigenetic modifications which are shown to influence chromatin structure and transcriptional regulation include DNA methylation, chromatin remodelling and histone modifications (Bernstein and Hake, 2006; Berger, 2007; Dolinoy, Weidman and Jirtle, 2007).

DNA methylation in mammals is a post replication modification which primarily involves the covalent addition of a methyl group (CH3) to the C-5 position of cytosine bases (Montero *et al*, 1992; Wolffe, Jones and Wade, 1999).

Since its discovery research has revealed the role of methylation in gene regulation and several biological processes including x-chromosome inactivation, genomic imprinting and ageing (Surani, 1998). It has already been established that aberrations in epigenetic mechanisms can give rise to several disorders such as cancer, autoimmune disease, imprinting disorders, neurological disorders and aging (Robertson and Wolffe, 2000; Jones and Baylin, 2002; Chen *et al*, 2010).

DNMT's are specialised DNA polymerases that catalyse cytosine methylation (Figure 30). The three DNMT's responsible for DNA methylation are DNMT1, 3a and 3b.

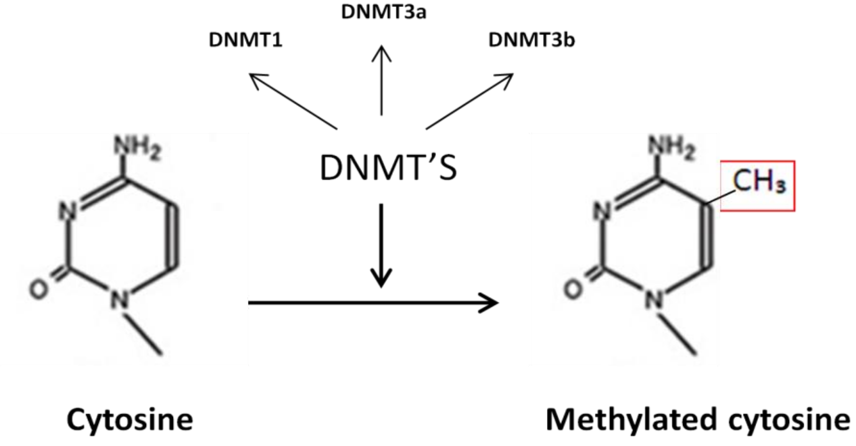


Figure 30. Cytosine methylation describes the addition of methyl group cytosine residues in DNA. Cytosine methylation is catalysed by DNA methyltransferase 1, 3a and 3b activity.

5.2.1.1 *Dnmt1*

Dnmt1 is ubiquitous in somatic cells and fundamental to the maintenance of DNA methylation patterns with each round of replication (Figure 31; Li, Bestor and Jaenisch, 1992; Lei *et al*, 1996; Robertson *et al*, 1999; Kanai *et al*, 2001). However, studies have reported that *Dnmt1* performs *de novo* methylation under carcinogenic conditions.

The catalytic activity of *Dnmt1* is carried out by the carboxy-terminal domain and the long N terminus regulatory region this allows the enzyme to distinguish between methylated and non-methylated DNA (Fang *et al*, 2001). The regulatory domain of *Dnmt1* also interacts with various cellular components including transcription factors (i.e. E2F1), tumour suppressor proteins (i.e. *Rb*) and histone deacetlylases (Robertson *et al*, 2000). The knockdown of *Dnmt1* results in genomic instability, impaired cell development and cell death (Lee *et al*, 2001; Brown and Robertson, 2007; Chen *et al*, 2007). The link between E2F mutation and genomic instability is helped by several proteins. Disruption of DNA binding activity between *Dnmt, Rb1* and E2F1 is associated with several cancers (Robertson *et al*, 2000).

Dnmt1 interacts closely with the pRb/E2F1 complex. Studies showed that the absence of pRB elevated levels of *Dnmt1* expression (McCabe, Davis and Day, 2005). This interaction between pRB, E2F and *Dnmt1* is facilitated by the existence of E2F binding sites in the *Dnmt1* promoter. Binding between E2F and its targets sites is sequence specific. Moreover *Dnmt1* can also influence the expression of genes involved in cell cycle control, DNA damage and repair and chromosomal instability (Brown and Robertson, 2007; Tan and Porter, 2009). This suggests that *Dnmt1* and E2F complexes produce DNA methylation effects in E2F binding sites (Robertson and Wolffe, 2000; Mortusewicz *et al*, 2005).

5.2.1.2 *Dnmt3a* and *Dnmt3b*

Dnmt3a and *Dnmt3b* were first identified in mammals and are responsible for *de novo* methylation and helps restore methylation patterns in imprinted genes (Figure 31; Okano *et al*, 1999). Extensive mouse studies have shown that deficiency of *Dnmt3a* and *3b* results in embryonic lethality and multiple organ death (Okano *et al*, 1999; Dodget *et al*, 2005). Studies have shown that increased *Dnmt3a* levels

promote polypopsis and may be involved in the development of several cancers including colon and HCC (Zhao *et al*, 2010).

Dnmt3a can suppress cell proliferation in HCC cell lines (Shafiei *et al*, 2008) and melanoma growth in mouse melanoma models (Deng *et al*, 2009). Zhao *et al* (2010) described the regulation of the TSG – PTEN and suppression of cell proliferation in *Dnmt3a* deleted cells. *Dnmt3b* is specifically required for the methylation of pericentric satellite repeats (Wienholz *et al*, 2010). Mutations in this gene is also detrimental and results in the development of the immunodeficiency disorder ICF (immunodeficiency, centromeric instability and facial anomalies). Shah *et al* (2010) found that truncated *Dnmt3b* isoforms commonly found in human tumours can disrupt embryonic development by inducing changes in DNA methylation. Other studies suggest that *Dnmt3b* interacts with *Dnmt1* to inactivate gene expression (Rhee *et al*, 2002; Leu *et al*, 2003).

CpG methylation exists in approximately 70% of the mammalian genome with the exception of CPG islands (Sadikovic *et al*, 2008; Illingworth *et al*, 2010). CpG islands are unmethylated GC rich domains ranging from 0.5 -5kb (Sadikovic *et al*, 2008). Although these regions account for only 1-2% of the mammalian genome, they have specific properties (Das and Singal, 2004). CpG islands provide epigenetic marks for the open chromatin (euchromatin) configuration which contains transcriptionally active genes (Bird, 1986; Illingworth *et al*, 2008).

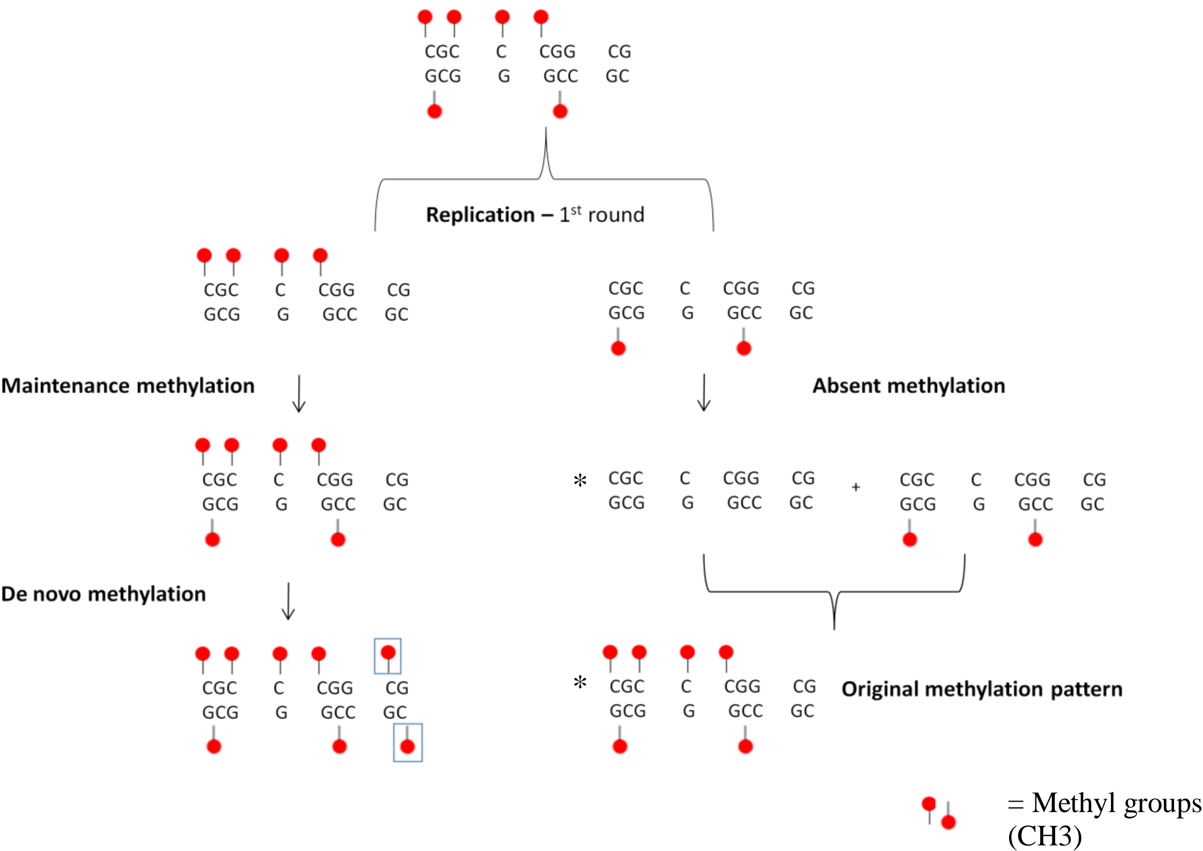


Figure 31. Inheritance of cytosine methylation patterns in DNA. Methylation patterns are crucial to gene expression and the maintenance of specific biological functions within a cell. Methylation patterns are passed onto daughter cells during mitosis. This enables methylation patterns to be maintained from cell to cell. However *de novo* methylation can disrupt the inheritance of DNA methylation patterns which exist in parent cells. Alternative DNA methylation mechanisms (*) can produce stable maintenance of methylation patterns even in the absence of Dnmt's. Unfortunately further insight is required to understand the mechanisms behind these alternative route. Methylated cytosines are represented as red dots.

5.2.1.3 Interaction between DNA methylation and chromatin

The interaction between DNA methylation and DNA-protein interaction plays a key role in chromatin structure, initiation of replication, and mismatch repair (Adams, 1990). However, DNA methylation may have a negative effect on DNA protein interactions as *Dnmt's* target sequences required for DNA cleavage, transcription factor binding and DNA protein binding (Watt and Molloy, 1988). This interaction between chromatin, DNA methylation may influence transcriptional regulation. Complexes consisting of binding proteins and methylated CpG's can stimulate the

recruitment of histone deacetylases (HDAC), *de novo* methylation and gene silencing (Figure 32; Bird and Wolfe, 1999; Billard *et al*, 2002; El-Osta, 2003). Histone modifying proteins such as HDAC promote heterochromatin configuration which is associated with gene silencing.

In addition histone methylations such as H3-K9 have been linked to pathological conditions such as cancer. Several studies have also reported the association between methylation of histone H3 at lysine 9 (H3-K9), heterochromatin formation and gene silencing (Oakes *et al*, 2007; Chen, Odenike and Rowley, 2010).

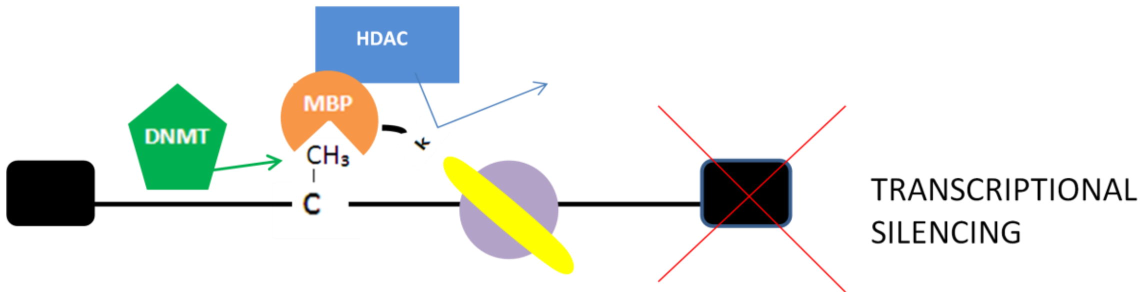


Figure 32. This model shows the association between Dnmt, HDAC's, methyl binding proteins and methylated cytosines. Aberrations in these complexes can influence chromatin and may result in the inactivation of adjacent cellular genes.

5.2.2 Detection of DNA methylation changes

Various methods are currently used to asses global and gene specific methylation including bisulphite reaction based methods (Frommer *et al*, 1992; Preuss and Pikard, 2007), chromatography (i.e HLPC), CPG island microarray (Khulan *et al*, 2006; Schumacher *et al*, 2006), TaqMan PCR (Kimura *et al*, 2009), Restriction Landmark Genomic Scanning (RLGS-M) and methyl sensitive restriction enzyme treatment (Bird and Southern, 1978). Bisulphite treatment is considered the gold standard of epigenetic studies.

Methylation abnormalities are subdivided into two categories: hypermethylation and hypomethylation.

Hypermethylation mostly occurs regionally within CPG islands and promoter regions. Hypermethylation is a major contributor to transcriptional repression (Bird and Wolffe, 1999).

Several studies have shown that genes affected by hypermethylation and gene silencing are implicated in several processes including cell cycle regulation, DNA repair, apoptosis, detoxification and drug resistance.

Gene specific methylation patterns have been identified in several cancers including esophageal squamous cell carcinoma (Yamashita *et al*, 2008), prostate cancer, HCC (Yang *et al*, 2003) and breast cancer (Feng *et al*, 2007). These cancers often present aberrant methylation within the vicinity of exon 1 or promoter regions of key tumour suppressor genes such as *P16ink4a* (Sharpless *et al*, 2002; Choi and Wu, 2005; Yang *et al*, 2005). Interestingly, ICF patients that exhibit abnormal DNA hypermethylation patterns fail to develop cancer (Tao *et al*, 2002; Ehrlich, Jackson and Weemaes, 2006). Hypomethylation occurs genome wide and affects highly and moderately repetitive DNA sequences present within transcription control regions, heterochromatin regions, endogenous retroviral elements or interspersed repeats (Takai *et al*, 2000; Ehrlich, 2002a,b; Saito *et al*, 2010).

Hypomethylation has also been observed in various cancers including HCC (Lapeyre and Becker 1979; Ehrlich *et al*, 1982; Lu *et al*, 1983; Bedford and van Helden, 1987). It is thought that hypomethylation and inactivation of genes involved in DNA damage repair and tumour suppression may trigger malignant transformation.

5.2.3 Methylation, mutagenesis and DNA damage

Several studies have suggested that key interactions exist between DNA damage homology, directed repair and DNA methylation (Cuozzo *et al*, 2007). Global deficiencies in methylated cytosines may activate DNA damage or create an environment which contributes to genomic instability and carcinogenesis. Lindahl, (1993) described the ability of cytosine methylation to increase the risk of mutagenesis.

Methylated cytosines interfere with DNA damage repair enzyme activity and may increase the risk of spontaneous point mutations originating from cytosines (Lindahl, 1993).

5.2.4 The role of viral integration in methylation

The introduction of foreign DNA into chromosomes can increase or reduce cellular methylation patterns and influence transcription of cellular genes (Remus *et al*, 1999; Heller *et al*, 1995). Provirus sequences such as the LTR, transgene and flanking genomic sequence may be subject to epigenetic mechanisms in the host genome (Harbers *et al*, 1981; Pannell and Ellis, 2001; Lavie *et al*, 2005). Non-integrating virus's such as HCV exhibits similar methylation patterns to the integrating HBV virus (Lee *et al*, 2003).Methylation is believed to have developed as a host genome defence mechanism against the uptake, integration or expression of foreign DNA (Karlin, Doerfler and Cardon, 1994; Tao and Robertson, 2003). Methylation can contribute to retroviral latency (Saggioro, Forino and Chieco-Bianchi, 1991; Finzi *et al*, 1997; Fang *et al*, 2001). Epigenetic mechanisms are also thought to have silenced the activity of endogenous retroviruses millions of years ago (Lavie *et al*, 2005).

The viral genome is usually methylated in the LTR to suppress viral gene expression and prevent viral replication close to genes flanking the proviral integration site (Challita and Kohn, 1994; Karlin *et al*, 1994). Challita *et al*, (1995) investigated the effect of RV vectors. Results showed that modification of the RV vector construct and inclusion of a demethylating fragment increased proviral expression and reduced DNA methylation levels following transduction of embryonic carcinoma cells (Challita *et al*, 1995).

The idea that viral integration is responsible for methylation changes in host genomes supports the concept of methylation as a host defence mechanism. However, viral induced methylation does not affect the ability of retroviruses to produce replication competent proviral DNA. Several studies have demonstrated the role of retroviruses and integration defective lentiviruses in methylation (Fang *et al*, 2001). Jahner and Jaenisch (1985) describe the ability of provirus integration to alter the methylation patterns of host DNA. Research shows that 20-25% of all retroviral integrations are found within 10kb of promoter elements and transcriptionally active genes involved in processes such as cell proliferation (Nienhuis, Dumbar and Sorrentino, 2006). This suggests that a high frequency of promoter methylation may be mediated by viral integration.

Integration site selection is dependent on various factors including GC content and distance of provirus insertion site from CpG islands (Hacker *et al*, 2006).

Hacker *et al,* (2006) found that LV and RV vectors favoured integration into regions within 10kb and 1kb of CPG islands respectively.

These studies suggest that RV and LV favoured sites are prone to epigenetic changes and play a role in gene expression and molecular signaling

Hacker *et al* (2006) also found a 1% increase in GC content could increase the probability of HIV-1 integration by 8.7%. In a separate study Kitamura, Lee and Coffin (1992) found that methylated CPG dinucleotides were favoured target sites of integration by avian leukosis virus. Other studies have shown that viral integration can induce *de novo* methylation within 1kb of the integration site. This suggests that epigenetic modification is an important factor in integration site selection.

The association between virus integration, DNA methylation and gene expression could present a serious danger to gene therapy patients treated with viral vectors.

Fang *et al* (2001) investigated the relationship between HIV-1 infection and methylation levels in genomic DNA. HIV-1 infection was shown to correlate with elevated *Dnmt1* levels and induce hypermethylation of cellular genes (including gamma interferon and *P16ink4a*) neighbouring proviral integration sites. This study also reported silencing of several genes. Tsai *et al*, (2002) also describe the ability of an oncogenic virus to stimulate hypermethylation and repress *e-cadherin* expression via activation of *Dnmt's*.

Hejnar *et al* (2003) investigated the role of RV provirus integration on DNA methylation patterns following transduction of Syrian hamster cells. This study reported a transient demethylation step that is followed by *de nova* methylation (Hejnar *et al*, 2003). In contrast Doerfler *et al*, (2001) described the existence of demethylation and *de novo* methylation at specific sites caused by viral transfection of murine cells.

The complex interaction between target site selection and viral mediated methylation and insertional mutagenesis suggests that several mechanisms involved in oncogeneis may also be deregulated by viral integration. This also supports the role of viral integration in oncogenesis (Doerfler *et al*, 1997 and 2001).

It is clear that understanding the epigenetic pathways in health and disease will facilitate novel approaches to the diagnosis and treatment of various diseases.

Genome wide studies were conducted to detect gene expression using microarray analysis.

It is clear that viruses exploit natural epigenetic mechanisms that can escape immune detection and increase their survival (Tao and Robertson, 2003). Viral integration influenced gene expression levels by interacting with E2F transcription factors and disrupting *Dnmt* activity.

The main hypothesis of this study was that LV infection induced methylation changes in the mouse genome. This study measured global methylation levels in EIAV, FIV and HIV samples infected *in vivo* and *in vitro*. In addition gene promoter methylation changes were measured via bisulphate sequencing and methylation specific PCR analysis. It was also hypothesised that *Dnmt* expression would be elevated in LV infected samples when compared to controls. *Dnmt* expression levels were measured in mouse liver tissue samples using *Dnmt* specific TaqMan probes.

Dnmt levels can directly or indirectly regulate the expression of E2F transcription factors and E2F target genes. It was hypothesised that changes in *Dnmt* levels may induce global dysregulation of E2F target genes. Microarray analysis was used to investigate the effect of LV integration on the expression of E2F target genes.

5.3.1 Microarray analysis of E2F targets *in vivo* to investigate the changes in E2F targets following LV integration

LV integration can greatly affect cellular mechanisms by altering the expression of host cellular genes. Analysis of differential expression of E2F target genes was carried out to determine the possible role of E2F transcription factors in vector mediated oncogenesis. The expression of E2F target genes was measured in EIAV associated tumours and compared to the expression of normal livers. EIAV infected liver tumours used in this study were found in Gerry, Adame and Jonas. Normal liver samples (n = 3) was also taken from Gerry, Adame and Jonas mice. Several E2F target genes without virus integrations were found differentially expressed in normal liver samples. To evaluate whether these changes could be associated with proviral infection tumour gene expression was compared to a spontaneous tumour liver that developed HCC at 568 days old. Spontaneous tumour liver was identified in a male mouse (mouse x). This mouse did not receive LV infection therefore it was used as an uninfected control for this study. In this analysis E2F 1, 2 and 3 transcription factors were differentially expressed in Gerry and Jonas normal samples when compared to when compared to normal liver (SpN) derived from mouse x. (Table 48). In addition, elevated levels of E2F 1 and 3 were found in Gerry normal liver when compared to SpN. In comparison Jonas normal liver showed a reduced level of E2F1 and E2F2 ($P < 0.05$) despite the elevated levels of the E2F 3 ($P < 0.05$).

Differential expression of E2F target genes in normal Adame, Gerry and Jonas liver samples was also obvious. Jonas showed the highest number of elevated E2F target genes (n = 31) followed by Gerry (n = 23) and then Adame (n = 13).

In contrast, only 11 E2F target genes were found to be upregulated in spontaneous tumour liver samples. All three E2F transcription factors (E2F1, 2 and 3) were upregulated in Gerry tumour samples. As E2F transcription factors are involved in activation of gene expression (such as tumour suppressor genes) it was interesting to see that 74 of the 94 E2F target genes were upregulated in Gerry tumour samples. Of these genes the tumour suppressors *Rb1* and *Trp53* are included. The expression of E2F3 was also increased in Adame and Jonas tumour samples. However E2F1 and E2F2 genes were downregulated in Jonas tumour samples (Table 48). *Prc1*, *Cd6*

and *Cdt1* are important genes that play a key role in cytokenesis, T cell activation and initiating DNA replication, respectively, and these genes can also exert oncogenic activity. Functional partners associated with *Prc1*, *Cd6* and *Cdt1* include DNA damage and repair genes, HCC target genes and *Dnmt* target genes. *Prc1*, *Cd6* and *Cdt1* were all upregulated in Gerry T1 and Jonas tumour samples. Interestingly *Prc1* was also increased in Adame and Jonas normal liver samples. This suggests that these genes are influenced by EIAV integration and may play a role in vector mediated oncognesis. Interestingly *Brac1* and *Brac2* were also differentially expressed in Gerry and Jonas tumour samples but failed to show significant changes in Gerry, Adame and Jonas normal livers (Table 48). This supports the involvement of these genes in carcinogenesis.

Bcl2 was previously shown by STRING and IPA analysis to belong to an oncogenic network that includes the EIAV insertion genes *Park7* and *Uvrag* (Figure 24). The increased expression of *Bcl2* in Gerry tumour (29.9%) and spontaneous tumour liver (10.3%) supports the importance of this gene in tumorigenesis and HCC development. However, results showed similar expression of *Bcl2* in Gerry normal liver (29.65) and Jonas normal liver (10.4%).

Pten is an important tumour suppressor gene and potent regulator of the P13K/Akt signalling pathway. Kim *et al* (2005) have demonstrated negative regulation of *Pten* by the *Park7* oncogene (Kim *et al*, 2005). QPCR analysis revealed reduced levels of *Park7*. Microarray analysis also revealed reduced levels of *Pten* in Gerry, Adame and Jonas tumour samples (Table 48). This suggests that viral integration disrupted the interaction between *Park7* and *Pten*.

Of the 94 E2F target genes analysed 77.7% were involved in DNA damage and repair mechanisms, 21.3% were oncogenes or known to exert oncogenic activity and 80.9% were categorized as HCC target genes.

5.3.1.1 E2F transcription factors and target genes are differentially expressed in EIAV infected tissues

No.	Gene symbol	Oncogene	DDR	E2F targets	HCC targets	DNMT targets	SpT vs SpN	G.N vs SpN	A.N vs SpN	J.N vs SpN	G.T vs SpN	A.T vs SpN	J.T vs SpN
1	*Jun*	●	●	●		●	-0.237	1.062	0.718	0.943	3.241	0.894	1.018
2	*Ccne1*	●	●	●	●	●	-0.014	0.003	-0.037	-0.010	0.632	-0.014	-0.029
3	*Ccne2*			●	●	●	0.047	0.038	0.074	0.012	-0.017	0.042	0.118
4	*Cdc25a*	●	●	●			-0.026	0.034	-0.061	-0.057	0.287	-0.054	-0.027
5	*Cdk2*	★	●	●	●	●	-0.101	0.373	-0.227	-0.179	0.287	-0.203	-0.219
6	*E2f1*	★	●	●	●	●	0.015	0.309	-0.042	-0.078	1.451	0.008	0.069
7	*E2f2*	★	●	●	●	●	-0.053	0.111	-0.073	-0.079	0.210	-0.031	-0.045
8	*E2f3*	★	●	●	●	●	-0.009	0.118	-0.045	0.028	0.123	0.097	0.064
9	*Tfdp1*		●	●	●	●	-0.083	-0.102	-0.021	-0.023	0.215	-0.036	-0.140
10	*Aurkb*		●	●	●		0.041	0.067	0.056	0.081	0.370	0.012	0.055
11	*Cdc20*			●		●	-0.057	-0.022	0.049	0.017	0.190	0.068	-0.103
12	*Prc1*	★	●	●	●	●	-0.053	0.095	0.155	0.174	2.472	0.135	0.064
13	*Cdc6*	★	●	●	●	●	-0.006	0.040	0.013	0.008	0.108	0.013	0.113
14	*Cdt1*	★	●	●	●	●	0.037	0.130	0.035	-0.008	1.086	-0.011	0.126
15	*Dck*			●	●	●	-0.021	-0.064	0.031	0.004	0.223	-0.001	-0.042

16	*Rrm2*		●	●	●	●	0.078	0.032	0.282	0.301	0.503	0.307	-0.407
17	*Tk1*		●	●			0.514	-0.342	0.812	0.634	1.552	0.559	0.185
18	*Top2a*		●	●	●		-0.041	-0.042	0.047	0.033	0.721	-0.008	-0.033
19	*Tyms*			●		●	-0.090	-0.139	0.057	0.122	0.827	0.128	-0.235
20	*Cdkn1c*			●	●		0.388	-0.141	0.077	0.070	-0.569	0.125	-0.116
21	*Cdkn2c*			●	●		0.015	-0.018	0.059	0.033	1.925	0.021	-0.078
22	*Cdkn2d*			●	●	●	0.033	0.051	-0.037	-0.035	0.143	-0.053	-0.070
23	*Rbl1*		●	●	●	●	-0.029	-0.031	-0.026	-0.016	0.526	0.016	-0.038
24	*Brca1*		●	●	●	●	0.016	0.049	0.040	0.013	-0.069	-0.017	0.115
25	*Brca2*		●	●	●	●	0.015	0.010	-0.046	-0.017	0.199	-0.035	0.039
26	*Bub1b*		●	●	●		-0.031	-0.016	0.004	0.004	0.243	-0.002	-0.005
27	*Bub3*			●			-0.022	-0.044	-0.095	-0.038	-0.032	-0.118	0.022
28	*Mad2l1*			●			0.002	0.044	0.027	0.052	0.202	0.025	0.029
29	*Cstf1*			●			0.168	0.360	-0.068	0.029	0.066	0.164	0.564
30	*Fen1*		●	●	●	●	-0.123	0.083	-0.109	-0.048	1.724	-0.049	-0.022
31	*Mgmt*		●	●	●	●	0.197	-0.012	-0.244	-0.018	-0.935	0.025	-0.191
32	*Mlh1*		●	●	●	●	-0.117	-0.077	-0.046	-0.161	0.026	-0.135	0.221
33	*Msh2*		●	●	●	●	0.042	-0.004	0.006	0.020	0.091	0.002	0.015
34	*Msh6*		●	●	●	●	-0.452	-0.395	-0.492	-0.622	0.737	-0.599	-0.276

35	***Pms2***		●	●	●	●	**-0.020**	**0.000**	**-0.050**	**-0.045**	**0.112**	**-0.059**	**0.024**
36	***Prkdc***		●	●	●	●	**-0.127**	**-0.068**	**-0.207**	**-0.122**	**0.129**	**-0.205**	**0.062**
37	***Rad51***		●	●	●	●	**0.029**	**0.018**	**0.046**	**0.051**	**0.159**	**0.027**	**0.022**
38	***Rad54l***		●	●			**0.033**	**0.107**	**-0.013**	**0.018**	**0.852**	**-0.010**	**0.081**
39	***Apaf1***		●	●	●	●	**0.002**	**-0.023**	**0.086**	**0.004**	**0.129**	**0.049**	**-0.030**
40	***Bad***		●	●	●		**0.109**	**0.062**	**-0.036**	**0.136**	**0.212**	**0.073**	**-0.166**
41	***Bak1***		●	●	●	●	**0.034**	**-0.052**	**-0.096**	**-0.084**	**-0.098**	**-0.076**	**0.057**
42	***Bcl2***	●	●	●	●	●	**0.103**	**0.296**	**0.035**	**0.104**	**0.299**	**0.063**	**-0.104**
43	***Bid***		●	●	●	●	**0.191**	**0.015**	**-0.029**	**-0.070**	**0.210**	**-0.055**	**0.153**
44	***Bok***		●	●			**0.026**	**0.834**	**0.051**	**0.114**	**1.657**	**0.169**	**1.049**
45	***Casp3***		●	●	●	●	**0.150**	**-0.009**	**0.072**	**-0.017**	**0.113**	**0.007**	**-0.130**
46	***Casp7***		●	●	●	●	**0.096**	**-0.371**	**-0.280**	**-0.223**	**-0.149**	**-0.268**	**-1.046**
47	***Casp8***		●	●	●	●	**0.039**	**0.197**	**0.063**	**0.012**	**0.898**	**0.106**	**-0.006**
48	***Map3k14***		●	●			**0.216**	**0.576**	**0.080**	**0.156**	**0.601**	**0.143**	**0.231**
49	***Map3k5***		●	●			**-0.041**	**-0.020**	**-0.011**	**-0.060**	**0.124**	**-0.024**	**-0.069**
50	***Eed***		●	●	●	●	**-0.033**	**0.223**	**-0.081**	**-0.058**	**0.061**	**-0.060**	**0.152**
51	***Fos***	●	●	●	●	●	**-0.028**	**0.305**	**0.663**	**0.964**	**0.304**	**0.694**	**0.440**
52	***Hey1***			●			**0.000**	**-0.012**	**0.000**	**0.013**	**0.195**	**0.006**	**-0.004**

53	*Hoxa5*	★	●	●	●	●	-0.049	-0.028	0.049	0.245	0.037	0.245	-0.035
54	*Fst*			●	●		-0.131	-0.104	0.030	-0.006	-0.309	-0.073	0.107
55	*Junb*	●	●	●	●	●	-0.171	0.299	0.139	0.767	1.278	0.334	1.110
56	*Mki67*			●	●	●	-0.181	-0.216	-0.211	-0.225	0.790	-0.179	-0.263
57	*Pold1*		●	●	●	●	0.047	0.144	-0.021	-0.088	1.404	-0.135	0.388
58	*Rad51ap1*		●	●			-0.032	0.021	0.009	0.002	0.091	0.027	0.069
59	*Rfc3*			●	●	●	0.018	0.038	-0.039	0.000	0.506	0.028	-0.034
60	*Rfc4*		●	●			0.051	0.295	-0.046	-0.057	0.782	-0.005	0.267
61	*Rrm1*		●	●	●	●	-0.071	-0.054	-0.024	0.001	0.125	0.011	-0.109
62	*Ccnb1*		●	●	●	●	-0.024	0.045	0.037	0.027	0.310	0.026	0.025
63	*Foxm1*		●	●	●	●	-0.029 Ns	-0.010	-0.026	-0.063	0.278	-0.021	-0.016
64	*H2afx*		●	●	●	●	-0.060	0.153	-0.226	-0.160	1.280	-0.080	0.073
65	*Tpx2*			●	●	●	-0.014	-0.002	0.000	0.001	0.183	-0.009	-0.019
66	*Uhrf1*	★	●	●	●	●	-0.011	0.160	-0.013	0.012	1.965	-0.004	0.085
67	*cdk1*		●	●	●	●	*	-0.018	0.053	0.044	0.872	0.110	-0.009
68	*H2afz*		●	●			*	0.771	0.131	0.290	1.842	0.218	0.491
69	*Mcm5*			●	●	●	0.035	0.263	0.235	0.240	3.045	0.258	0.022
70	*Lig1*		●	●	●	●	-0.034	0.151	0.058	0.196	1.739	-0.024	0.133
71	*Mcm2*		●	●	●	●	-0.003	0.016	-0.037	-0.016	2.021	-0.088	-0.088

72	*Mcm6*			●	●	●	0.060	0.088	0.128	0.161	2.725	0.056	0.035
73	*Pcna*		●	●	●	●	-0.022	-0.002	0.020	-0.012	0.011	-0.021	0.051
74	*Tacc3*		●	●	●	●	-0.008	-0.027	-0.016	-0.008	-0.013	-0.029	-0.039
75	*Dmap1*		●	●	●	●	0.016	0.091	-0.243	-0.225	0.054	-0.182	0.352
76	*Mat1a*			●		●	0.081	-0.066	-0.358	-0.408	-0.985	-0.535	-0.178
77	*Hdac2*		●	●	●	●	0.031	0.065	-0.058	-0.078	0.135	-0.040	-0.002
78	*Daxx*		●	●	●	●	0.019	0.084	-0.083	-0.026	0.121	-0.101	0.266
79	*Mat2a*		●	●	●	●	0.239	0.405	0.124	0.081	0.108	0.159	0.304
80	*Trdmt1*			●		●	0.101	-0.025	-0.022	-0.034	0.013	-0.029	0.011
81	*Dnmt1*	★	●	●	●	●	0.022	-0.884	-0.702	-0.762	0.317	-0.016	0.022
82	*Dnmt3b*	★	●	●	●	●	-1.021	-0.884	-0.702	-0.762	-0.812	-0.715	-0.560
83	*Cyclin D1*	●	●	●	●	●	-0.209	-0.451	-0.270	-0.267	-0.198	-0.239	-0.683
84	*Cdkn2a*		●	●	●	●	0.054	0.006	0.025	0.033	0.083	0.083	0.029
85	*Igf2r*			●	●		-0.003	-0.215	-0.502	-0.338	-0.341	-0.510	-0.319
86	*Igf2*	★	●	●	●		-0.038	-0.126	0.359	0.393	-0.084	0.313	2.388
87	*Smad2*		●	●	●	●	0.149	0.025	0.005	-0.033	0.192	0.097	-0.140
88	*Pten*		●	●	●	●	-0.054	-0.085	-0.042	-0.093	-0.017	-0.048	-0.112
89	*Ctnnb1*	●	●	●	●	●	-0.274	-0.502	-0.148	-0.416	-0.774	-0.327	-0.789

90	***Socs1***						**-0.046**	**-0.021**	**-0.153**	**-0.120**	**-0.253**	**-0.124**	**0.162**
91	***Gstp1***						**-1.171**	**1.629**	**0.625**	**0.613**	**1.071**	**0.760**	**1.083**
92	***Cdkn2b***						**-0.131**	**-0.080**	**-0.166**	**-0.138**	**-0.027**	**-0.159**	**-0.068**
93	***Rb1***						**-0.039**	**0.097**	**0.124**	**-0.004**	**0.169**	**0.019**	**-0.136**
94	***Trp53***						**-0.004**	**0.065**	**-0.247**	**-0.191**	**1.050**	**-0.156**	**-0.096**

Table 48. Changes in the expression of E2F target genes in EIAV infected liver and tumour samples. Spt = Spontaneous tumour liver. SpN = normal liver derived from mouse X. G.N, A.N, J.N represent Gerry normal, Adame normal and Jonas normal tissue samples respectively. G.T, A.T and J.T represents Gerry tumour, Adame tumour and Jonas tumour tissue samples respectively. Gene expression data recorded are all relative to spontaneous tumour liver samples. Fold changes in gene expression (logFC) were calculated and statistical analysis performed on raw data. Values annotated in red font represents gene expression levels with statistically significant P. Values (<0.05). ★ is used to represent genes that are not classified as oncogenes but known to possess oncogenic activity. HCC targets and DNMT target genes include known functional partner genes. Genes not included in the microarray are marked with an asterix (*).

5.3.2 Changes in DNA *methylase* expression in LV infected tissues

Microarray data showed altered expression of *Dnmt1* and *Dnmt3b* in the EIAV infected liver samples included in this study (n = 3). A marked reduction in the *de nova* methyltransferase enzyme - *Dnmt3b* was found in Gerry, Adame and Jonas normal liver samples. In comparison the maintenance methyl transferase *Dnmt1* was found to be down regulated in EIAV infected normal livers (n = 3) and 2 out of 3 liver tumours (Table 48).

DNA methylation is dependant on *Dnmt* activity to initiate, maintain and establish DNA methylation. This chapter presents data to support a possible link between viral integration and *Dnmt* expression in the mouse samples infected *in utero* with EIAV, FIV and HIV vectors. To accurately evaluate gene expression, QPCR was used to detect changes in *Dnmt* levels using gender matched liver samples (uninfected) as calibrator samples with the ΔΔCT method of gene expression analysis. Results showed elevated levels of *Dnmt*1 in Fiona normal liver ($P<0.001$) and all 5 FIV infected tumour samples ($P<0.05$) when compared to uninfected livers (Table 49).

The existence of *de novo* methylation was demonstrated by the observation of elevated levels of *Dnmt3a* in FIV treated normal liver samples ($P<0.001$) and in 3 out of 5 liver tumour samples ($P<0.01$). This suggests that *Dnmt* genes and methylation patterns are altered as a result of FIV integration. *Dnmt3b* levels were also measured in FIV infected liver samples and elevated levels were found in all normal livers ($P<0.05$) and in 4 out of 5 liver tumours ($P<0.001$; Table 49 and Figure 33 and 34). These results correlate with the expression levels of *Dnmt1* and *Dnmt3a* expression found using microarray and supports the involvement of LV infection in epigenetic mechanisms that are known involved in oncogenesis.

Sample name	*Dnmt1* (+/-SEM)	*Dnmt3a* (+/-SEM)	*Dnmt3b* (+/-SEM)
FIV normal			
Fiona N	15.98	8.17	53.05
Flora N	0.71	8.46	3.24
FIV tumour			
Fiona T1	24.81 (0.19)	73.52 (0.12)	30.77 (0.31)
Fiona T2	61.26 (0.23)	16.11 (0.14)	7.68 (0.15)*
Fanny T1	1.79 (0.49)	0.89 (0.06)	8.89 (0.18)
Fanny T2	3.42 (0.14)*	4.35 (0.11)	47.38
Frank T2	1.96 (0.29)	0.83 (0.18)	0.13 (0.16)*

Table 49. *Dnmt* expression in mouse tissue samples after FIV infection.
(*) Is used to denote that the P. Value obtained were above 0.05 confidence levels. All P.Values obtained were relative to a gender matched uninfected liver sample. *Dnmt* levels are described as relative methylation levels ($2^{\Delta\Delta CT}$) calculated by ΔΔCT analysis.

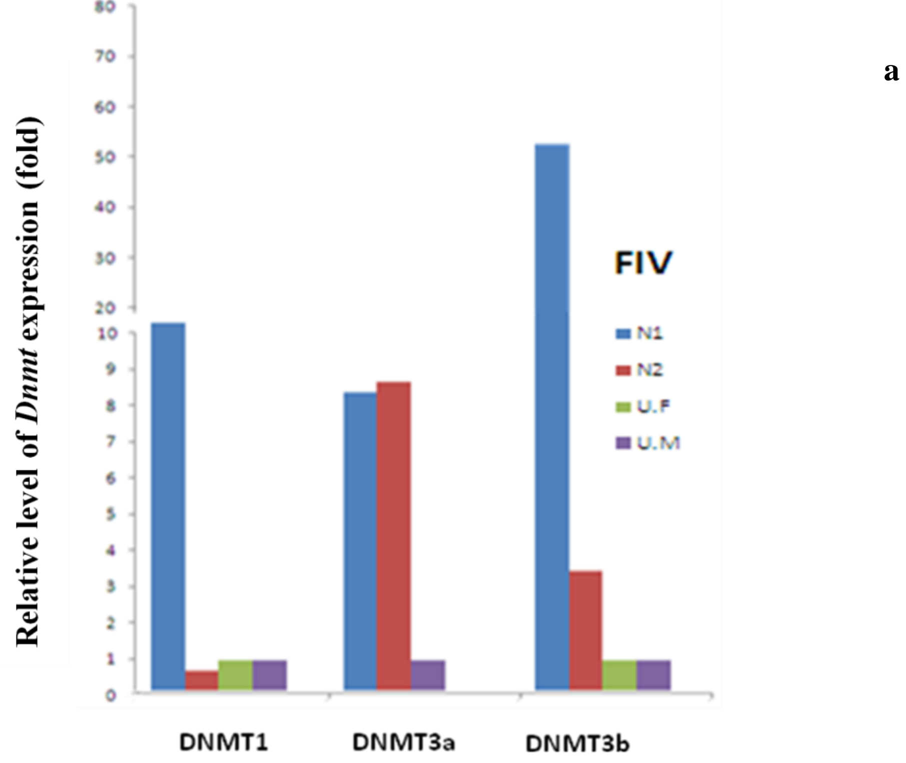


Figure 33. Histogram representing *Dnmt* levels in FIV infected samples. Each bar represents $2^{\Delta\Delta CT}$ values produced by FIV infected normal liver samples compared to uninfected livers. A) N1 & N2 represents Fiona normal and Flora normal liver samples respectively. UF and UM represents Uninfected female and male liver samples respectively.

Because aberrant methylation patterns are often associated with cancer the expression of *Dnmt* gene expression levels were evaluated in tumour samples associated with FIV infection (Figure 34 and table 49). Fiona T1 and T2 samples showed consistently higher levels of *Dnmt* expression when compared to uninfected male and other tumour samples (Figure 34).

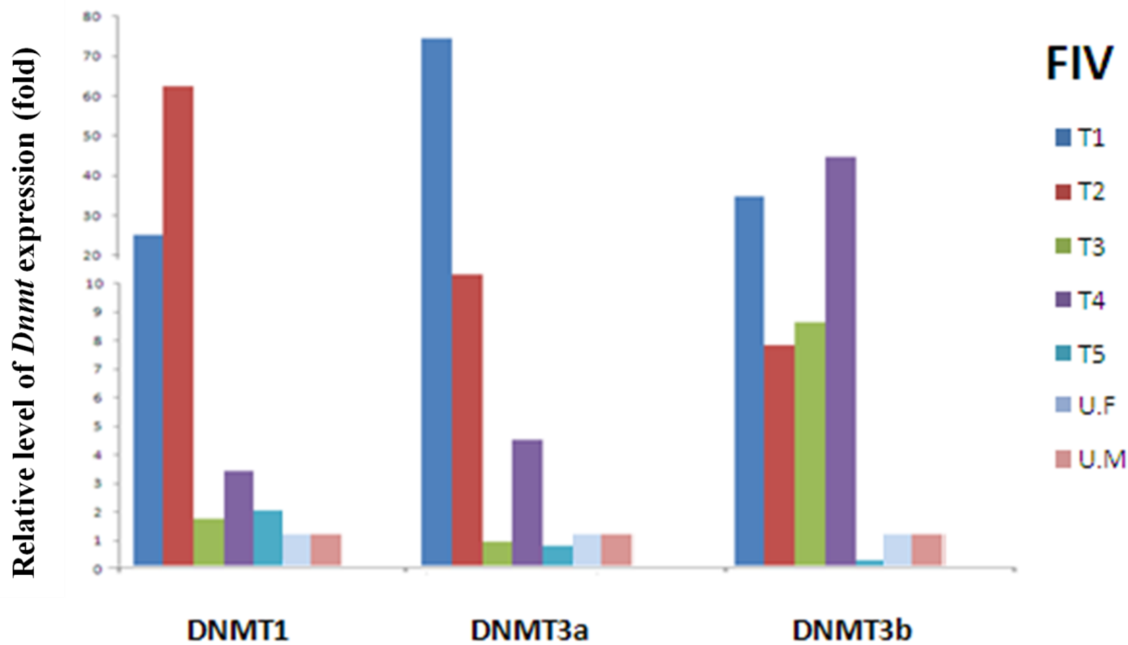


Figure 34. Histogram representing the relative gene expression values of *Dnmts* in tumour samples infected with the FIV vector. T1-T5 represents Fiona T1 and T2, Fanny T1 and T2 and Frank T2 tumour samples described in Table 49. UF and UM represents Uninfected female and male liver samples respectively.

QPCR analysis found that *Dnmt1* was not expressed in Ikea and Gerty normal liver samples (Figure 35). However both Ikea and Gerty showed elevated expression of *Dnmt3a* and *Dnm3b* as shown in Figure 35. Analysis of EIAV infected tumour samples showed elevated levels of *Dnmt1*, *Dnmt3a* and *Dnmt3b* as shown in Table 50 and Figure 36. As HIV infected mice did not develop tumours *Dnmt* expression could only be examined in normal liver samples. QPCR analysis showed elevated expression of *Dnmt1* and *Dnmt3b* in HIV infected normal livers (Figure 37 and Table 51).

Sample name	*Dnmt1* (+/-SEM)	*Dnmt3a* (+/-SEM)	*Dnmt3b* (+/-SEM)
EIAV normal			
Ikea normal liver	0.15 (0.08)	21.36 (0.20)	3.36
Gerry normal liver	0.71(0.20)	30.17 (0.26)	2.96 (0.51)
EIAV tumour			
Adame T1	685.63 (0.21)	50.56 (0.18)	30.77 (0.31)
Gerry T1	61.26 (0.23)	19.20 (0.26)	65.64 (0)

Table 50. *Dnmt* gene expression data obtained from EIAV infected mice.

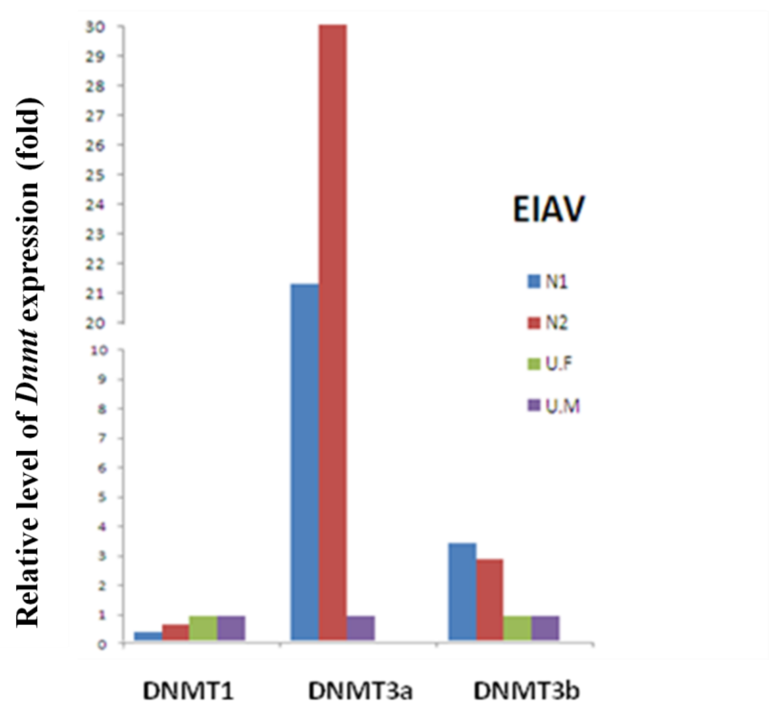


Figure 35. *Dnmt* gene expression data obtained from EIAV infected mice. N1 and N2 represent Ikea and Gerry normal liver samples. UF and UM represents Uninfected female and male liver samples respectively.

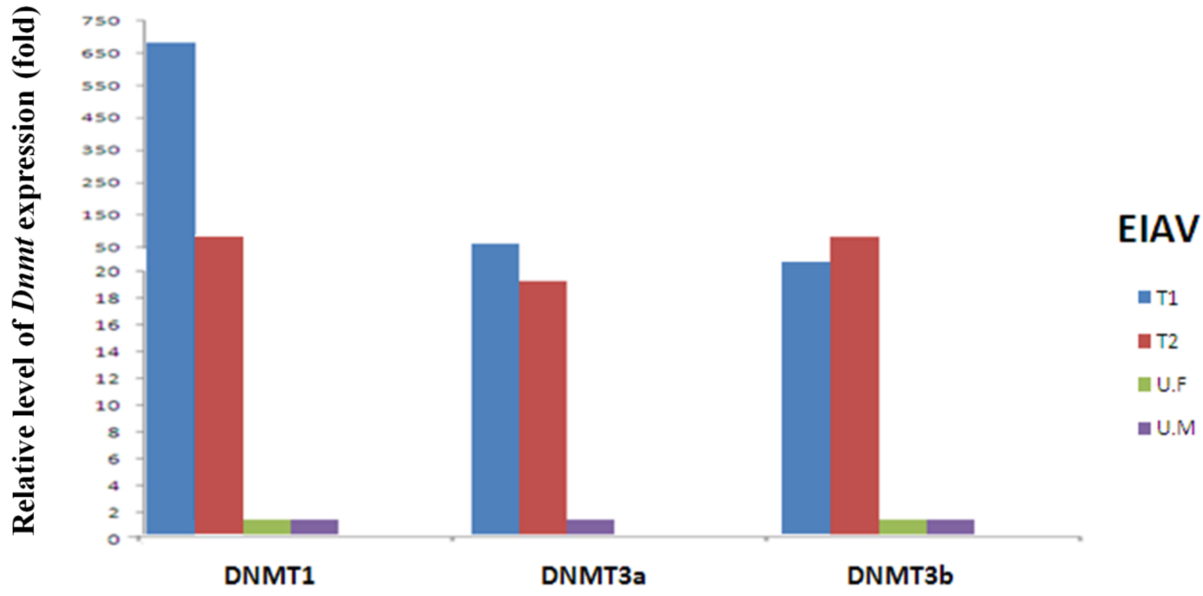


Figure 36. *Dnmt* gene expression data obtained from EIAV infected mice.

T1 and T2 represent Adame tumour 1 and Gerry tumour 1 samples, respectively. UF and UM represents Uninfected female and male liver samples respectively.

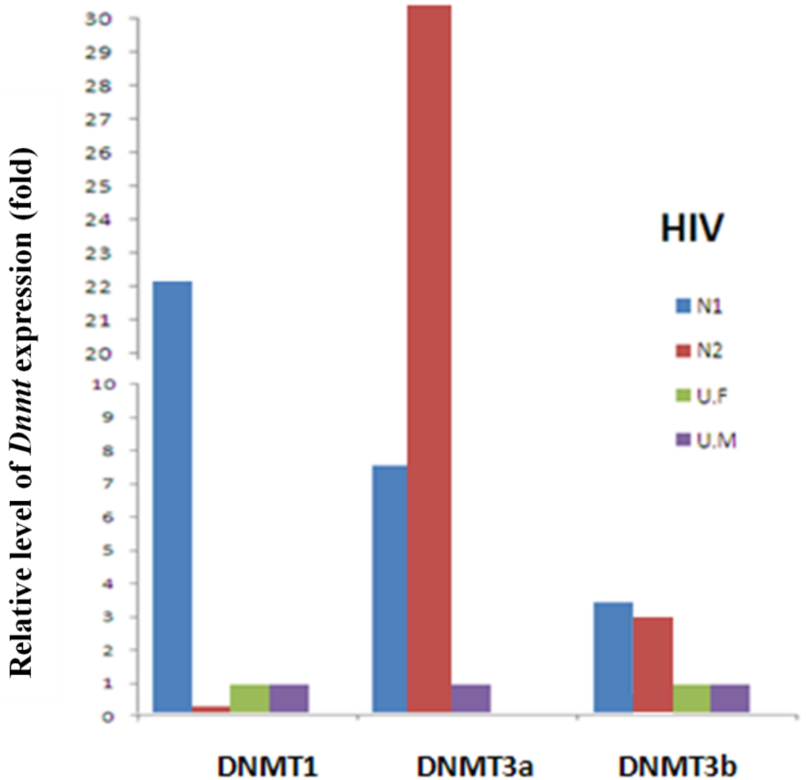


Figure 37. *Dnmt* gene expression data obtained from HIV infected normal liver samples N1 and N2 represents Tiggy and Bruce liver samples respectively. UF and UM represents Uninfected female and male liver samples respectively.

Sample name	*Dnmt1* (+/-SEM)	*Dnmt3a* (+/-SEM)	*Dnmt3b* (+/-SEM)
HIV			
Tiggy	22.15 (0.16)	7.21 (1.50)	3.36 (0.80)
Bruce	0.06 (0.45)	30.17 (0.08)	2.96 (0.30)

Table 51. Expression of *Dnmt* genes in HIV infected samples.

To investigate whether DNA methylation was increased globally in the mouse genome methylation levels were measured in LV infected samples. A global methylation kit was used to detect methylated cytosines present in DNA samples. Global changes in methylation levels are shown in Figures 38 and 39. Average levels in methylation are also shown. Increases in global methylation were seen in normal livers obtained from EIAV (44.84%); FIV (60.73%) and HIV (100%) infected mice respectively (Figure 38a). Global DNA methylation was also examined in the LV associated tumours of EIAV (n=5) and FIV (n=7) including an FIV associated ovarian cyst. Global methylation levels in these tumours increased to 90% and 110% for EIAV and FIV, respectively, compared to no increases in methylation found in uninfected controls (Figure 39). These results suggest DNA methylation levels may be influenced by LV infection and could represent a potentially independent mechanism of genotoxicity leading to oncogenesis.

5.3.3 Observations of global methylations associated with *in utero* application of LV vectors

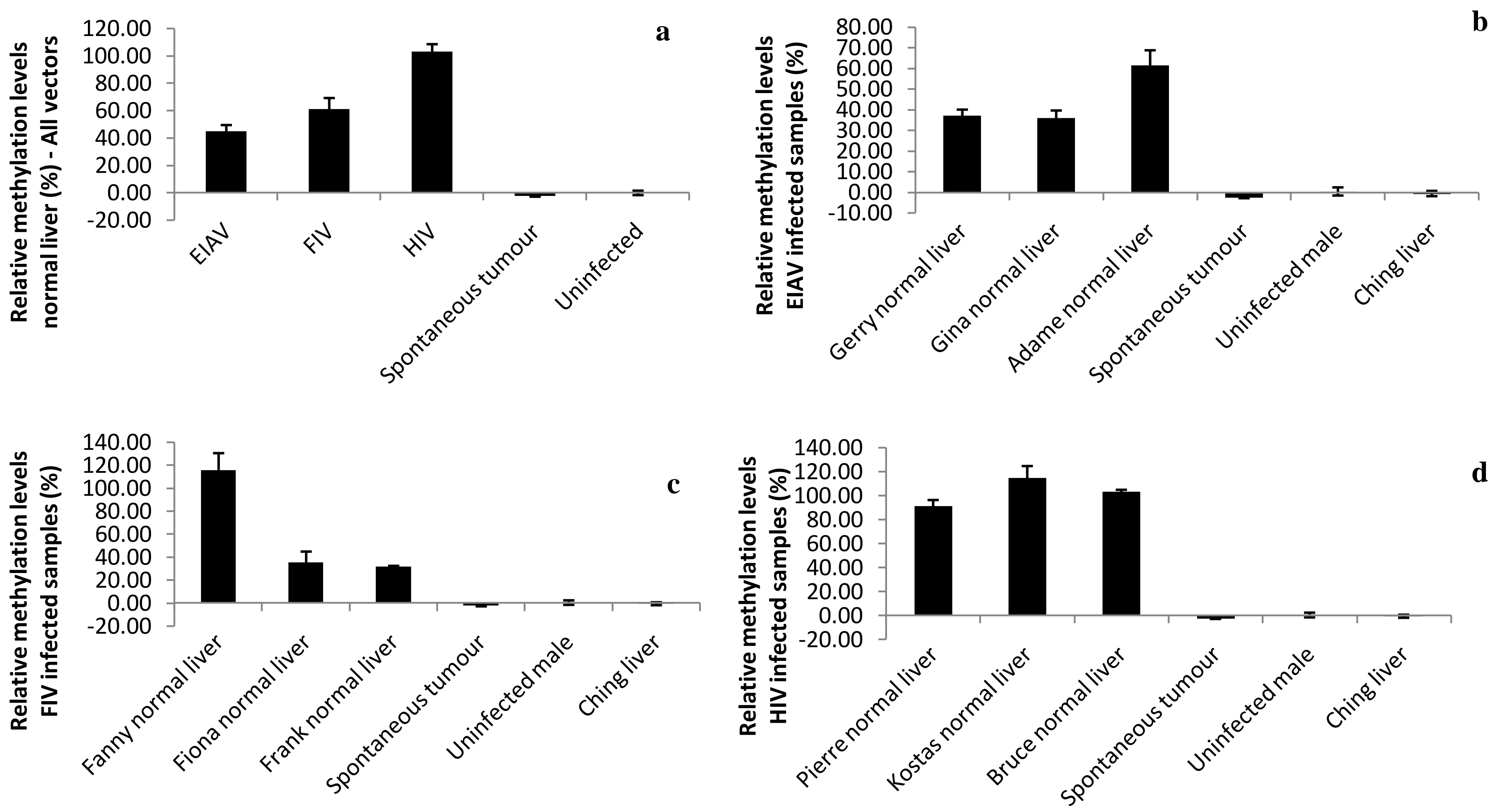


Figure 38. DNA methylation levels in normal tissues are associated with LV transduction. Global methylation status was compared between normal hepatic tissue (a-d) and a control group of a spontaneous tumour sample and male and female uninfected liver samples. Ching liver was extracted from a female mouse that did not receive LV infection. Relative methylation values (%) were calculated from average absorbance values. Error bars represent the standard error of means derived from triplicate assays. Global methylation was significantly higher in EIAV (b) and HIV (d) samples (<0.001). FIV infected samples also showed an increase in global methylation (<0.05).

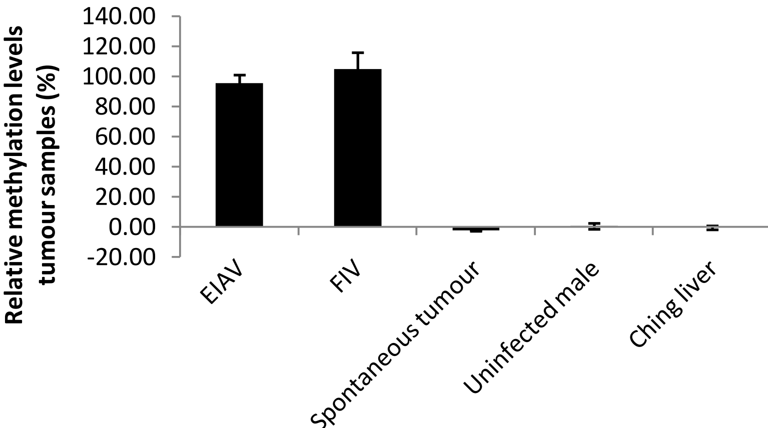


Figure 39. DNA methylation changes in hepatocellular carcinoma tissues transduced with LV vectors.

Relative methylation levels are represented as black bars in the graph. Error bars represent standard deviations between triplicate absorbance values. The global methylation status of FIV infected (1-7) and EIAV infected tumour samples (1-5) varied. The control groups are a spontaneous tumour and two uninfected liver samples (Uninfected male and Ching liver). These were found without elevated levels of methylation.

5.3.4 Measurement of the effects of LV infection on methylation *in vitro*

To determine whether LV infection increased DNA methylation at early time points after infection *Dnmt* expression and global methylation levels were investigated *in vitro* directly after LV infection of the HepG2 liver cell line. If this was to occur this would support the data presented from *in vivo* analysis and would suggest that LV integration or infection influences the epigenetic regulation of gene expression and could contribute to oncogenesis.

HepG2 cells were infected and *methylase* expression and levels of methylation examined at 3 time points (6 hours, 24 hours and 30 hours) post infection by EIAV, FIV, HIV and MLV vectors. To determine whether changes in methylation were associated only with integration 2 integrase negative HIV vectors were used for infection that contained either the CMV or SFFV promoters driving eGFP expression. DNA and RNA were harvested from infected and uninfected tissue samples immediately after each timepoint. Global methylation assays performed on EIAV vectors carrying the CMV promoter showed a small increase in methylation when compared to uninfected control DNA samples extracted at this point (6 hrs, 24hrs and 30 hrs). These samples showed increased methylation levels in all LV and RV treated cells (excluding EIAV at 6hrs). In all LV and RV treated cells (excluding EIAV – 6hrs) small levels of DNA methylation was seen when compared to uninfected cells taken at 24 hours (2.67) and 30 hrs (5.00). Several LV infected samples showed high global methylation levels. These included HIV integrase negative (CMV and SFFV), HIV, FIV and MLV infected HepG2 cells (Table 52; Figure 40).

5.3.4.1 Global methylation levels: *In vitro* study

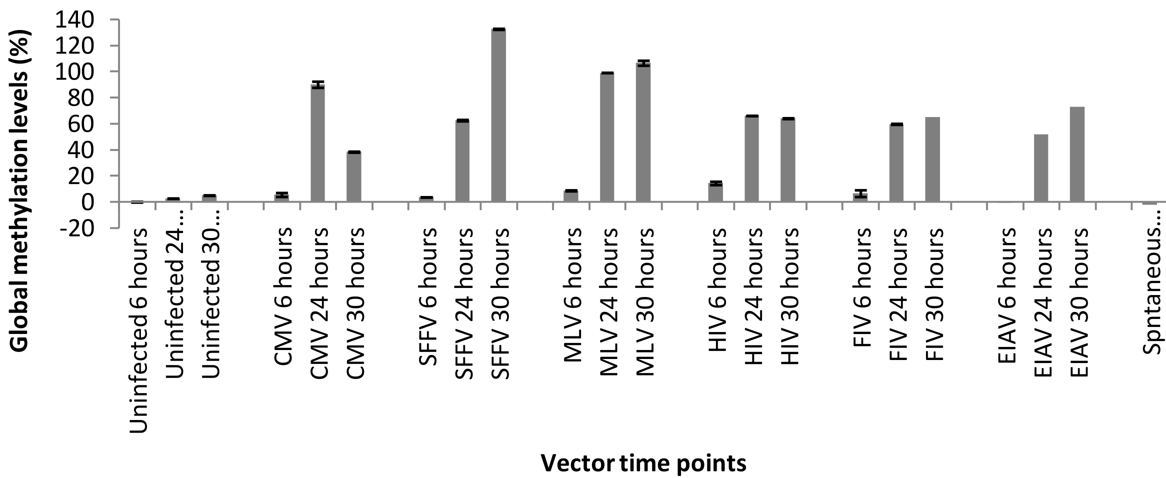


Figure 40. Global methylation levels in uninfected, LV treated and RV treated HepG2 cells at 6hrs, 24hrs and 30 hrs post infection. Absorbance readings at 450nm were used to detect methylated cytosines. Standard error bars represent SEM values obtained from triplicate absorbance readings.

Global methylation levels: *In vitro* study

Sample	Time point (hrs)	Relative methylated levels (%) SEM (%)
Controls *		
Uninfected (blank)*	6 hours	0 (0.30)
	24 hours	2.67 (0.13)
	30 hours	5.00 (0.20)
CMV*	6 hours	5.50 (0.59)
	24 hours	90.00 (1.50)
	30 hours	38.33 (2.36)
SFFV*	6 hours	3.67 (0.38)
	24 hours	62.50 (1.45)
	30 hours	132.50 (0.05)
MLV	6 hours	8.67 (0.55)
	24 hours	99.00 (0.40)
	30 hours	106.50 (3.65)
HIV	6 hours	14.33 (0.37)
	24 hours	66.00 (0.10)
	30 hours	64.00 (1.90)
FIV	6 hours	6.50 (0.75)
	24 hours	59.67 (1.30)
	30 hours	65.00 (0.065)
EIAV	6 hours	-0.67 (0.30)
	24 hours	51.67 (0.87)
	30 hours	72.67 (2.58)
Spontaneous tumour*	*	-2.50 (0.39)

Table 52. Global methylation levels. DNA methylation levels measured at early time points (6, 24 and 30hrs) in uninfected or post LV/RV infection (integrase negative, MLV, FIV. HIV and EIAV). * is used to represent uninfected and HIV integrase negative vector infected samples. These control samples were used as references of normal global methylations levels.

5.3.4.2 *Dnmt* methylation levels: *In vitro* study

QPCR showed low levels of *Dnmt3a* expression in HepG2 cells infected with the integrase negative vector (CMV promoter), FIV, HIV (P <0.05) and EIAV (P <0.001) 6 hours after LV infection (Table 53). These results were compared to data obtained from uninfected HepG2 cells (6 hrs post infection). Results suggest that *de novo* methylation did occur at early time points (6 hrs) after LV infection but at very low levels. *Dnmt3a* levels also appeared to peak 30 hrs after treatment with FIV (P<0.001) and EIAV (P<0.01) vectors. SFFV infected cells also showed higher levels of *Dnmt3a* expression between 6 and 24hrs after infection.

Interestingly HIV and MLV infected cells showed unique patterns of *Dnmt* transcription levels. MLV infected cells showed a rise in *Dnmt3a* expression at 6 hours after infection. This rise was followed by a 98.74% reduction in transcription levels from 26.0 at 6hrs to 0.32 at 30 hrs post infection. HIV infected cells also demonstrated a rise in *Dnmt3a* expression at 6hrs post infection (P<0.05) and again at 30 hrs post infection (P<0.001) (Table 53; Figure 41).

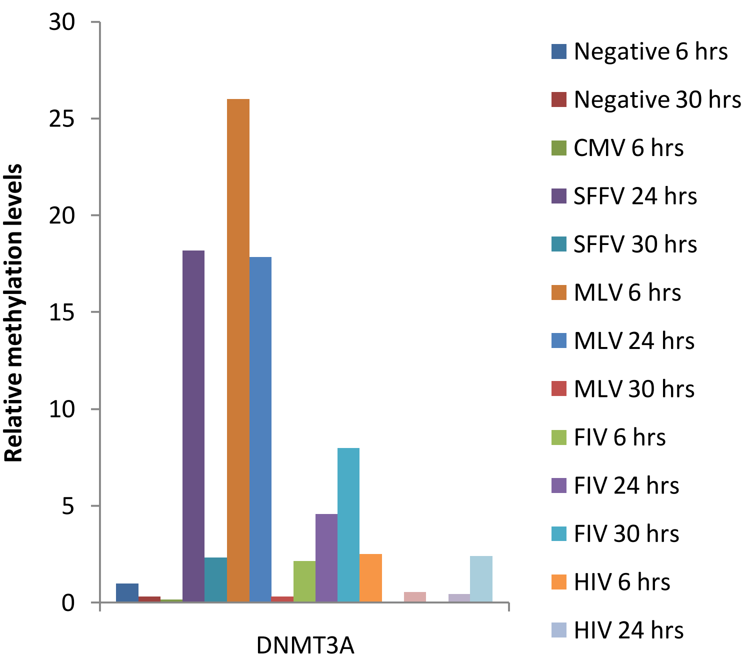


Figure 41. *Dnmt3a* levels seen in negative (uninfected) and vector treated (MLV, FIV, HIV and EIAV) HepG2 cells. CMV – Cytomegalovirus & SFFV = Spleen Focus-Forming Virus Integrase negative vectors carrying CMV or SFFV vectors are described as CMV or SFFV respectively. QPCR analysis failed to detect *Dnmt1* and *Dnmt3b* expression in infected HepG2 cells.

Sample name	*Dnmt3a*	P values N6 Compared to n6	P values V6
Negative 6 hrs	1	*	*
Negative 30 hrs	0.316	*	*
Integrase negative (CMV) 6 hrs	0.175	NS	*
Integrase negative (SFFV) 30 hrs	2.342	<0.01	*
MLV 6 hrs	25.987	<0.05	*
MLV 24 hrs	17.85	<0.01	NS
MLV 30 hrs	0.321	<0.001	<0.01
FIV 6 hrs	2.134	NS	*
FIV 24 hrs	4.608	<0.001	<0.01
FIV 30 hrs	8.007	<0.001	<0.05
HIV 6 hrs	2.506	<0.05	*
HIV 24 hrs	0.07	<0.001	<0.001
HIV 30 hrs	0.556	<0.001	<0.05
EIAV 6 hrs	0	<0.001	*
EIAV 24 hrs	0.463	NS	<0.001
EIAV 30 hrs	2.408	<0.01	<0.001

Table 53. *Dnmt* gene expression data represented as relative methylation levels ($2^{\Delta\Delta CT}$). Student Test analysis was used to determine P. values when compared to negative 6 and vector matched CT values for *Dnmt3a* at 6hrs post infection (V6). NS was used to describe P. Values above significant levels (<0.05). Values marked with an asterix (*) were excluded from study due to poor *18sRNA* readings.

5.3.5 Analysis of VCN following the delivery of LV vectors to liver cell *in vitro*

Sample name	Time point	Copy Number per Diploid (SEM)
Uninfected	6 hours	0.00 (0)
	24 hours	0.35 (0.35)
	30 hours	1.94 (0.32)
CMV	6 hours	482.86 (0.25)
	24 hours	16458.86 (0.58)
	30 hours	6561.14 (0.18)
SFFV	6 hours	11.76 (0.05)
	24 hours	2132.46 (0.36)
	30 hours	578.91 (0.02)
HIV	6 hours	1733.40 (0.02)
	24 hours	6834.20 (0.53)
	30 hours	2405.51 (0.01)
FIV	6 hours	147.74 (0.45)
	24 hours	898.37 (0.24)
	30 hours	6054.66 (0.16)
EIAV	6 hours	20.70 (0.08)
	24 hours	77.70 (0.09)
	30 hours	58.96 (0.13)

Table 54. VCN data obtained using QPCR analysis.

VCN analysis was measured in genomic DNA extracted from HepG2 cells treated with LV and RV. Results showed an increase in VCN derived from cells collected at 6 to 30 hrs post infection (Table 54). Most VCN values were above 12. QPCR failed to detect viral sequences in uninfected cells as expected.

6.1 Oncogenesis associated with differential gene expression with *in utero* and *in vitro* application of LV vectors

Haematopoietic stem cells (HSC) infected with RV vectors have been successfully used to treat several immunodeficiencies such as ADA-SCID (Hacein-Bey-Abina *et al*, 2002; Gaspar *et al*, 2004). However, an early French gene therapy clinical trial conducted by Cavazzana-Calvo and Fischer reported leukaemia development in patients that received gene modified HSC's (Cavazzana-Calvo *et al*, 2005). The trial involved transduction of autologous CD34+ bone marrow cells with RV vectors containing the γC gene that were transduced *ex vivo* and returned to each patient. Vector insertions in the vicinity of the *LMO2* oncogene induced spontaneous T cell leukaemia (Nienhuis *et al* 2006).

Since these observations were made in HSC's these cells have been used as models of leukemogenesis and several studies have sought to understand the molecular mechanisms of vector mediated genotoxicity (Montini *et al*, 2006). However comparisons between solid tumours and malignant haematopoiesis are limited by differences in these molecular mechanisms and oncogenic networks underlying these pathogenic disorders. This makes it difficult to estimate the potential risk of vector mediated genotoxicity from results obtained in malignant haematopoiesis.

It is hope that the development of genotoxicity models will improve the safety of gene therapy using non-viral and viral vectors (Von Kalle *et al*, 2004). Despite the advantages of using HSC's for gene therapy somatic gene transfer could offer genetic modification for various cell types *in situ*. Thus far, however, models to establish a link between LV and RV integration and its associated side effects following somatic gene therapy have not been created and with the evidence that genotoxicity may be influenced by gene expression the potential for vector associated genotoxicity following *in vivo* gene transfer is, as yet, unknown (Kustikova *et al*, 2010).

Extensive retroviral insertion site mapping has been used to reveal vector insertion site preferences some of which include integrations into well known protoncogenes such as *Ccdn2* and *Bmi1* (Qasim *et al*, 2009; Hacein-Bey-Abina, 2008). Highlighting

these insertion preferences is useful to predict the likelihood of vector mediated oncogenesis and therefore in this study insertion site analysis was carried out.

The concept of fetal (*in utero*) gene therapy is based on the prevention of genetic conditions before birth and to achieve permanent correction of such diseases by stable transduction of cell populations before they manifest in the unborn child (Coutelle *et al,* 2005).

The application of somatic gene transfer to the foetus is relatively new to the field of gene therapy. RV vectors prefer to insert into actively transcribing genes (Hacker *et al*, 2006; Ambrosi *et al*, 2008; Persons, 2010), however, in the foetus several genes that are known oncogenes are present in this state (switched on), this means in the foetus the potential for IM is greater. Key genes involved in processes such as cell cycle metabolism and differentiation are also highly expressed in the foetus and therefore it is important to know if these genes are also capable of contributing to oncogenesis if they become mutated after integration. Indeed, the potential for carcinogenesis induced by IM of viral vectors *in utero* highlights the usefulness of the foetus as a potentially useful genotoxicity model.

A previous study conducted by Themis *et al* (2005) investigated the correlation between LV integration and vector mediated insertional mutagenesis using the foetal mouse. EIAV (SMART2 EIAV) and HIV (HR'SIN-cPPT-S-FIX-W) vector constructs were used to treat haemophiliac MF-1 mice *in utero* at 16 days gestation. These experiments demonstrated a relationship between the insertion sites of the EIAV vector and the development of HCC (Themis *et al*, 2005) with 80-90% of foetal EIAV treated mice developing clonally derived HCC. In the same study mice treated with HIV based vectors were not found to develop tumours. These results demonstrated the foetal mouse to be a useful *in vivo* system that can be used for the analysis of vector safety.

The oncogenesis observed in LV treated mice was believed, initially to be caused by the inclusion of the woodchuck post translation regulatory element (WPRE) in the vector construct. The WPRE element has been included in viral vectors in several studies to improve transgene expression and the performance of viral vectors in gene transfer (Zufferey *et al*, 1999). In the wild-type configuration the WPRE domain consists of an X gene (coding for an 140 amino acid peptide) and has been shown to

play a role in HCC (Zufferey *et al*, 1999; Kingsman *et al*, 2005; Higashimoto, *et al*, 2007; Schambach *et al*, 2007). In gene therapy vectors such as HR'SIN-cPPT-S-FIX-W a truncated form of the X gene is present, however, the promoter and start codon for the X gene is mutated to prevent X expression. Hence, this was believed to be the reason why mice treated with HR'SIN-cPPT-S-FIX-W failed to develop tumours. In the EIAV vector these mutations do not exist and therefore it was X gene expression that was suspected to be the cause of oncogenesis.

In the Themis laboratory the role of the X protein in oncogenesis was further investigated through the use of non primate FIV based vectors (pLIONhAATGFP) and modified EIAV vectors (SMART2ΔW) that also included the truncated X gene with a mutated promoter and start codon. Mice treated with pLIONhAATGFP and SMART2ΔW induced clonal HCC but HR'SIN-cPPT-S-FIX-W treated mice once again failed to show signs of tumour development. Mice treated with these LV vectors (excluding HIV) developed tumours at a frequency of 4/6 (SMART2 unmodified), 4/10 (SMART2 modified vector) and 4/8 (pLIONhAATGFP). These results implicated the role of LV integration site selection in vector mediated oncogenesis. Research described in this thesis focused on retrieving LV integration sites (within a window of 100kb of the proviral insertion) in FIV and EIAV infected samples. Integration sites retrieved from foetal mice were used to identify common insertion sites (CIS) and predict the risk of vector mediated oncogenesis associated with *in vivo* gene transfer.

Proviral insertions sites were retrieved from EIAV and FIV infected mouse samples using LAM PCR retrieval and 454 or Automated DNA (www.dnaseq.co.uk) sequencing to provide sequence data for BLAST analysis using the NCBI database. This identified RefSeq genes with vector integrations and gave information for vector integration profiling. The identity of these genes supports the possibility that LV integration may be associated with HCC. The EIAV vector appeared to integrate into genes associated with cellular proliferation and cancer and several oncogenes were found with EIAV insertions. FIV also showed a preference for integration into genes associated with cell growth and proliferation.

To investigate whether insertional mutagenesis was implicated in oncogenesis the expression of EIAV and FIV insertion genes were examined to detect alterations in

their expression. RT PCR analysis was first used to show the trends in expression for FIV insertion genes and this data was validated using QPCR analysis. The expression of EIAV insertion genes retrieved was quantified using QPCR analysis and microarray analysis.

High density microarray analysis is often used to validate and confirm gene expression data obtained from QPCR analysis. However, QPCR analysis is often believed to yield a more accurate measure of changes in gene expression (Dorak, 2006; Pallas *et al*, 2008). This increased sensitivity is particularly useful to studies that measure transcription levels in heterogeneous cell populations as observed in the HCCs found in this study.

By identifying genes altered in expression with EIAV LV vector insertions and relating these to genes in pathways associated with these insertion genes bioinformatic analysis was able to reveal potential oncogenic pathways that could be assigned to HCC and the vector insertion site (s). STRING and IPA networks were therefore most valuable for this and provided evidence suggestive of insertional mutagenesis being the cause of HCC in the treated mice. These networks included genes such as *Tnfrsf19*, *Pscd3*, *Park7*, *Uvrag*, *Pah*, *Mrlp23* and *Bre*.

Correlations between two SMART2 insertions identified in Jonas large tumour were investigated. These genes included the *Uvrag* and the oncogene *Park7*. The latter gene plays a key role in oxidative stress responses (Clements *et al*, 2006; Vasseur *et al*, 2009). Several studies have reported that overxpression of *Park7* in cancer may facilitate the ability of tumour cells to sustain cellular proliferation (Clements *et al*, 2006; Sitaram *et al*, 2009; Vasseur *et al*, 2009). Interestingly both *Uvrag* and *Park7* were retrieved from the same tumour sample in a previous study. QPCR analysis showed that the SMART2 insertion into intron 12 of the *Uvrag* gene resulted in up regulation of the tumour suppressor gene *Uvrag* by 128%. Microarray analysis also revealed an increase in *Uvrag* by 78%. This contradicts studies that show suppressed expression of *Uvrag* in liver metastasis (Goi *et al*, 2003). In contrast *Park7* was found to reduced in QPCR analysis. Both *Uvrag* and *Park7* form part of an oncogeneic network that includes key cancer associated genes including *Trp53*, *Bax*,

and *Bcl2*. Interestingly, all of these genes were also shown to be deregulated in Jonas tumour samples by microarray analysis.

In Jonas large tumour *Bre* (Brain and reproductive organ-expressed protein) was shown to be downregulated when compared to gender matched normal and tumour control samples. This antiapoptotic gene shows multiple roles in several cellular processes including cellular differentiation and regulation of the inflammatory cytokine TNF-alpha through its interaction with the TNF receptor 1 (microarray *Tnf1a*). Several studies have described the overexpression of *Bre* in HCC (Chan *et al*, 2008; 2010). *Bre* has also been described as a RIS hit (n = 2) in the RTCGD database. Interestingly several cancer associated genes were found in the *bre* protein-protein network. These genes include *Smad* (1 and 4) and *Brac* (*1* and *2*) proteins. Although *Smad 1* and *4* were not found in the microarray analysis data for *Smad2* was included in this study. Levels of *Smad2* expression were shown to be downregulated by microarray analysis. *Bre* is a component of the BRCA1-A complex that specifically recognises histones H2A and H2AX at DNA lesion sites. This interaction between BRAC1-A and H2A/H2AX histones recruits BRACA1-BARDI heterodimers to sites of DNA damage at double stranded breaks. This implicates *Bre* in DNA damage response and repair pathways. Indeed, this pathway to HCC was supported by increased levels of *Brac1* (11.5%) and *Brac2* (3.9%) in the Jonas tumour sample. The levels of *Brac1* and *Brac2* were not seen in Adame and Gerry tumour samples. However, the results of this study showed that EIAV integration into *Bre* reduced its expression and possibly disrupted stress responses and led to a reduction in *Smad2* and elevated levels of *Brac1* and *2*. *Mark3* (Map/microtubule affinity-regulating kinase) was also retrieved from Jonas large tumour samples where it was found to be increased. *Mark3* plays a role in phosphorylation of microtubule associated proteins, ATP binding, nucleotide binding, and protein serine/threonine kinase activity. *Mark3* share a close homology with *Markl1* that is often ovexpressed in HCC. Integration of the EIAV provirus into *Katna1* in Jonas large tumour resulted in the reduction of this gene. However an association between *Katna1* and cancer associated genes was not found by bioinformatics. Other SMART2 inserted genes that were shown upregulated in Jonas large tumour included *Mrlp23* and *Rabgef1*. However, these genes were also

increased by greater levels in the Jonas small tumour sample. Similar expression of *Rabgef1* shown by microarray analysis supported the role of the *Ras* signalling in this tumour development. The implication of these findings in Jonas large tumour suggests that cooperation between EIAV insertions may occur.

In the Adame tumour sample QPCR analysis revealed upregulation of *Acvr2a* (Activin type II receptor) following an EIAV integration into this gene. However, microarray did not reveal differential expression of this gene. Activins belong to a family of important signalling proteins known as transforming growth factor-beta (*TGFβ*) (Phillips *et al*, 2001; Chen *et al*, 2006). *TGFβ* possess several functions and are thought to repress tumour progression through its inhibition of the proto oncogene *c-Myc* and *hTERT* (Li *et al*, 2006). Activins can also suppress cancer cell growth in early cancer development (Tsuchida *et al*, 2009). Signalling for Activin and TGFβ occur through different serine/threoinine kinase receptor subunits. However, both pathways share common activation of Smad proteins (Phillips *et al*, 2001). Deregulation of TGFB and activin signalling has been implicated in cancer and metastasis (Li *et al*, 2006; Tsuchida *et al*, 2009).

In the same tumour the expression of the *Pah* gene was found downregulated. *Pah* catalyses the catabolism of the essential amino acid phenylalanine. Aberrant regulation of *Pah* activity results in the metabolic disorder phenylketonuria. Interactions between *Pah* and several genes including *Hnf4a*, *Slco1a4* and *Pcbd1* were found using STRING analysis. Interestingly, *Slco1a4* and *Pcbd1* were found differentially expressed in the microarray data of Adame tumour.

Hnf4a is related to *Hnf1a* and known to be important functional partners that regulate the expression of the *Pah* gene. QPCR analysis revealed downregulation of both *Hnf4a* and *Hnf1a* genes by 55% and 62% respectively. This correlates with findings that describe reduced *Hnf4a* levels in HCC and liver cancer (Lazarevich and Alpern, 2008). Furthermore, *Hnf4a* was reduced in microarray analysis of *Hnf4a* by 53% in this tumour.

Networks relating insertion genes to cancer genes were not only found in the EIAV infected tumours. In Fiona T2 pLIONhAATGFP FIV insertions were found in intron 1 of the *Cyth3* gene. *Cyth3* also known as *Pscd3* was shown to be significantly

upregulated by RT PCR and QPCR analysis when compared to a gender matched tumour without a proviral insertion into this gene. *Pscd3* is involved in activation of the P13K/AKT signalling pathway (Hafner *et al*, 2006; Kolanus, 2007). Aberrations of the P13K/AKT signalling pathway is thought to play a critical role in cancer development (Fuss *et al*, 2006; Hafner *et al*, 2006). *Arf* can induce *p53* and interact with *Ras* and *Myc* genes to suppress cellular proliferation. Moreover aberrations in *Arf* activity are implicated in the activation of cancer (Hashimoto *et al*, 2004). STRING analysis provided strong evidence for the role the pLIONhAATGFP insertion into *Pscd3* and oncogenesis. *Pscd3* was found to interact with several cancer associated genes via an oncogeneic network that included *Akt1*, *Trp53* and *Pten*.

Tnfrsf19 was retrieved from Flora Ovarian Cyst samples. The expression of *Tnfrsf19* was completely silenced following integration of pLIONhAATGFP upstream of this gene. Tumour necrosis factor-α receptors are capable of activating necrosis (via caspase-independant cell death mechanisms) which results in oxidative stress. However, aberrant expression of *Tnfrsf19* may implicate this gene in the escape of cancer cells from cell death. Moreover, overexpression of *Tnfrsf19* can activate JNK signalling. Studies show that JNK signalling may be activated by the non-canonical Wnt signalling pathway which is already implicated in tumour development (Weston and Davies, 2002). Therefore, aberrant expression of JNK signalling induced by Wnt signalling may be a possible mechanism of oncogenesis in these mice.

Gene expression studies failed to find an association between *Nek9* (Frank T1) and *Coro7* (Frank T2) insertions and cancer associated genes.

In conclusion, therefore, from these analyses no common insertion sites (CISs) could be found. However, several pathways to tumour development were obvious. These data suggest pathways to HCC are complex and that several pathways may exist for solid tumour development. Many more mice would therefore need to be treated before CISs are discovered using the foetal mouse model.

Montini *et al* (2009) showed that VCN can contribute to vector mediated oncogenesis. By investigating the role of VCN in HCC development in EIAV, FIV and HIV treated mice it was hoped that increasing numbers of vector copies may

correlate with the frequency of oncogenesis or at least tumours would be expected to have similar copy numbers since each mouse received the same dose of vector. Southern blot analysis was used to establish the copy number of each tumour and this was used as a standard value for VCN analysis by QPCR of LV genomes in tumours and infected normal livers. VCN values varied widely regardless of the dose of vector used for infection. Several EIAV and FIV infected samples presented with above 10 LV copies in normal and tumour samples. The highest VCN values were obtained in Gerry normal (45.16) and tumour samples (86.34 and 202.13) which were also supported by Southern blot data. Hence, it was concluded that VCN was difficult to control in the foetal mouse.

In comparison, VCNs in HIV treated mice were all below 3. This could explain the why no tumours were identified in HIV treated mice.

Because CISs were not identified in the LV associated tumours and in many cases only a single oncogene could be identified with a provirus insertion, alternative mechanisms of vector mediated genotoxicity were suspected. Once of these mechanisms was host epigenetics involving methylation. The introduction of foreign DNA into chromosomes can increase or reduce cellular methylation patterns and influence transcription of cellular genes (Heller *et al*, 1995; Remus *et al*, 1999). Moreover provirus sequences such as the LTR, transgene and flanking genomic sequence may be subject to epigenetic mechanisms in the host genome (Harbers *et al*, 1981; Pannell and Ellis, 2001; Lavie *et al*, 2005). Indeed alterations in methylation can also contribute to cancer development (Robertson and Wolffe, 2000; Jones and Baylin, 2002). Therefore this study investigated the role of LV infection in liver cancer development by using *in vivo* and *in vitro* studies on methylation levels in tumours and normal infected livers.

Firstly, global methylation levels were investigated as an indicator that host methylation had occurred following infection. Relative methylation levels for EIAV and FIV treated tumour samples were 95.43% & 104.79% respectively. In comparison uninfected control samples produced no changes in relative methylation levels. Despite the similarities in average methylation levels obtained from EIAV and FIV samples it was the Jonas large tumour that demonstrated the highest level of DNA methylation relative to uninfected liver and normal EIAV infected liver

samples. This may have contributed to the silencing of *Park7* as discussed earlier via promoter methylation. In a previous study on this tumour (unpublished data) the role of promoter methylation of the cancer associated gene *Hmgcs2* was measured and found not only that this gene was methylated but also that *Hmgcs2* was down regulated in expression. Interestingly, *Hmgcs2* is known downregulated in over 90% of *c-Myc* associated HCCs (Camarero *et al*, 2006; Ralph *et al*, 2010). Hypermethylation was found at 2 CpG islands located in the *Hmgcs2* promoter. This hypermethylation was observed in Jonas normal samples whereas hypomethylation of *Hmgcs2* was evident in the Jonas tumour sample. This is in line with hypermethylation followed by hypomethylation in specific genes and cancer. Moreover the altered methylation levels found in the *Hmgcs2* promoter correlated with the downregulation of *Hmgcs2* expression measured by QPCR of this gene.
The global methylation analysis in normal liver samples also demonstrated increased relative methylation levels of 44.84% (EIAV), 60.73% (FIV) and 103.04% (HIV) respectively. This provided strong evidence for host defence against LV infection by epigenetic means. This study investigated the effect of LV infection on epigenetic (methylation) changes in the host genome. However, additional factors such as gender, age, environment differences (between experimental and control) or sickness in mouse samples may have contributed epigenetic changes observed. This effect could be investigated in future studies.

DNA methyl transferases (*Dnmts*) play a key role in *de novo* methylation (*Dnmt3a* and *3b*) and the maintenance (*Dnmt1*) of methylation patterns in the mammalian genome. It was therefore important to show that the observed changes in global methylation found in infected livers and tumours correlated with changes in *Dnmt* expression. Although tumour analysis was included in this study it was not assumed that methylation changes in these tumours could be clearly assigned to LV infection since hypermethylation followed by hypomethylation is a common hallmark of cancer and may not have been vector associated. However, QPCR analysis was used to measure the expression of *Dnmt 1*, *3a* and *3b* in clonally derived mouse tumours and their respective normal samples. A marked reduction in *Dnmt1* was found in EIAV infected normal liver in contrast to no change in the levels of *Dnmt1* in uninfected livers. However increased levels of *Dnmts3a* and *3b* were found in both

EIAV tumour and infected normal liver samples. A similar trend in *Dnmt* expression was shown in FIV infected samples suggesting this change could be related to LV infection and not only as a result of oncogenesis. Microarray data also showed altered expression of *Dnmt1* and *3b* in all EIAV infected normal liver samples included in this study (n = 3). A marked reduction in the *de novo* methyltransferase enzyme – *Dnmt3b* was found in Gerry, Adame and Jonas liver samples. In comparison the maintenance methyl transferase *Dnmt1* was found to be down regulated in only 1 EIAV infected normal liver samples (n = 3) and 1 out of 3 liver tumour samples (Table 48).
QPCR results showed elevated levels of *Dnmt expression* in FIV and HV infected liver samples in tumour samples and their respective normal livers when compared to uninfected liver.

The evidence found for involvement of DNA methylation associated with LV infection and possibly tumour development required that this study be repeated *in vitro*. To study whether changes in DNA methylation could be identified at early time points after LV/RV infection HepG2 cells were infected with each several LV and a retroviral vector (RV). In addition, to determine whether methylation changes were associated only with integration HIV integrase minus vectors carrying either the CMV or SFFV promoters driving gene expression were used.
Global rises in DNA methylation occurred after 6 hours however, these rises were small in comparison to the increased levels of methylation found at 24 and 30 hours after LV and RV infection (Table 52). All levels of global DNA methylation were compared to results obtained from uninfected HepG2 cells at 6hrs where no increase was found. Interestingly, the greatest increases in global methylation changes were seen in cells infected with MLV and the CMV driven HIV integrase negative vector suggesting methylation was induced by infection and not only due to integration. HIV infected cells also showed consistently increasing levels of global DNA methylation at 24 and 30 hours post infection. Again levels of global methylation in uninfected control sample were unchanged. The measurement of *Dnmt* levels in HepG2 infected cells concurred with global increases in methylation. *Dnmt3a* levels were mainly altered in expression following LV and RV infection and these increases were found highest at 24 and 30 hours post infection in MLV and HIV

infected cells. Importantly, since the *Dnmt3a* methylase establishes *de novo* methylation and this was found elevated rather than the *Dnmt1 methylase* that is responsible for maintaining methylation patterns this innate cellular response is most likely due to infection and is probably a mechanism aimed at silencing the vector. To strengthen the results of these studies similar experiments could be conducted in primary liver cells.
Evidence of LV high VCN was shown by QPCR of vector specific sequences. This was especially true for non integrating vectors.

Because coordination exists between E2F regulation and methylation it was suspected that the increased global methylation could ultimately cause alteration in the expression of this transcription factor and its target genes. E2F binding sites in target genes contain CpG regions that also serve as targets for regulation by *Dnmts* (Campanero, Armstrong and Flemington, 2000). Other studies have also demonstrated that *Dnmt3b* target genes are also controlled by E2Fs such as E2F6 (Velasco *et al*, 2010). Interestingly, E2F6 has also been shown to interact with repressive chromatin-modifying enzymes that play a role in methylation and epigenetic regulation of transcription. Moreover, *Dnmts* may interact with the regulation of several E2F family members and this could contribute to several cancer related changes in gene expression where oncogenes and tumour suppressor genes are involved (McCabe *et al*, 2005; 2006).

94 E2F target genes were found disregulated by microarray analysis of tumours and infected normal livers of these mice. Of the 94 E2F target genes analysed 77.7% were involved in DNA damage and repair mechanisms, 21.3% were known oncogenes or shown to exert oncogenic activity and 80.9% were categorised as HCC target genes. Microarray analysis also showed that E2F 1, 2 and 3 transcription factors were differentially expressed in Gerry and Jonas clonally derived tumours and their respective normal livers. Jonas showed the highest number of elevated E2F target genes in normal liver samples (n = 31) followed by Gerry (n = 23) and Adame (n = 13). In contrast, only 11 E2F target genes were found to be upregulated in spontaneous tumour liver samples.

As E2F transcription factors are involved in the activation of gene expression (such as tumour suppressor genes) it was interesting to see that 74 of the 94 E2F target genes found in Gerry tumour liver included important cancer genes that are known to be associated with *Rb1* and *Trp53, Brac1* and *Brac2*. The expression of E2F1 and 3 was also increased in Adame and Jonas tumour samples. However E2F2 expression was significantly downregulated in Jonas tumour samples (Table 48).
These findings suggest that infection by proviruses may in some way cause these changes in E2F or its target gene expression and this may be via the methylation route. Several E2F targets are known to be involved in oncogenesis (Müller *et al, 2001*; Nevins, 2001). Therefore viral integration into E2F target may be an indirect mechanism of vector mediated oncogenesis. Indeed, this may act in associated with insertional mutagenesis leading to HCC.

Future studies will further explore lentiviral integration and its contribution to gene expression changes within the host genome in more control and infected mice. The generation of this data will significantly enhance the statistical power to detect changes in gene expression and methylation. The role of LV insertions and its involvement in cancer as shown by bioinformatics analysis could be further explored by using SiRNA to target cancer associated genes in zenograft models.

Taken together the findings of this study demonstrate that LV vectors may mediate genotoxicity either directly by causing changes in gene expression of inserted genes or indirectly by causing methylation of important cellular genes following an innate immune response to the vector and this could lead to HCC. The fact that the HIV vector did not associate with HCC may be due to this vector not being mutagenic to inserted genes. This vector did, however, associate with changes in global methylation levels and could over time contribute to clonal expansion of infected cells and tumour development. Since no current long-term model exists to test this hypothesis this remains to be elucidated.

The fact that both the EIAV and FIV vectors used in this study are found to cause HCC in the fetally treated adult mice means these vectors, at least in the mouse, are genotoxic. The fact remains that all vectors used in this study alter methylation in the HepG2 human cell line and they should be modified to reduce this effect. This may be done using optimised vector sequences with reduce CpG sites used for methylation or by identifying what elements of these vector are causing the innate immune response. As for mutagenesis, thus far the cause of differential gene expression of the genes with insertions of these vectors has not been elucidated here and this requires investigation if these vectors are to be used confidently for gene therapy without side effects.

APPENDIX 1

EIAV

- SMART2hFIX
- SMART2Z
- SMART2ZWΔ

$1.2x10^7$

MF1 mice

FIV

- pLIONhAATGFP

$1x10^7$

MF1 mice

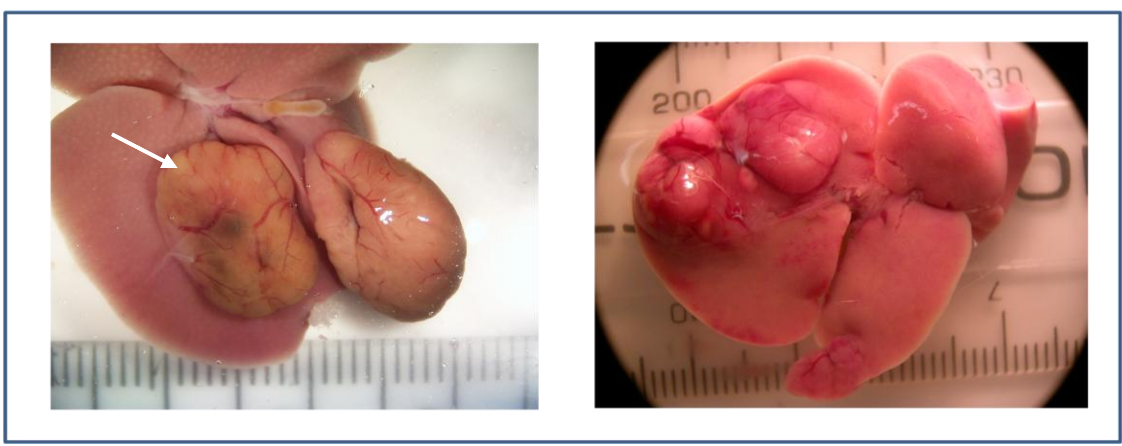
HIV

- HR'SIN-cPPT-S-FIX-W

$1.2x10^7$

MF1 mice

Lentiviral transduction of mice: a) Graphical image representing the dose and vector constructs used to treat MF-1 mice. EIAV and FIV treated mice received titers of approximately $1.2x10^7$. Despite this both vectors generated highly variable VCN's in infected mouse samples. b) The observations seen in this study included the development of liver tumours and the development of an ovarian cyst in one mouse.



APPENDIX 2

Theoretical sequence of pLIONII hAAT GFP
(FIV vector genome plasmid)

**CTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTAT
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGA
ACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGT
TTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC
GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAA
GCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCT
CCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTA
ACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGC
CTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTA
CCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTG
GTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTT
TGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGG
TCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTA
AATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTG
AGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCG
TGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGC
GAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCG
AGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGG
AAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAG
GCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGAT
CAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTC
CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGC
ATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAA
CCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATAC
GGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTT
CGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTC
GTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAA
CAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCA
TACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT
ACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAA
AAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGC
GTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACA
TGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCC
GTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAG
AGCAGATTGTACTGAGAGTGCACCATAAAATTGTAAACGTTAATATTTTGTTAAAATTCG
CGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCC
CTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGA
GTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG
ATGGCCCACTACGTGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAG
CACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGTAA
GCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC
TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGT
CAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCT
TTTGCTGGCCTTTTGCTCACATGGCTCGACAGATCTTCAATATTGGCCATTAGCCATATT**

ATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACGTTGTATCT

ATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGGCATTG
ATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATAT
GGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCC
CCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCA
TTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA
TCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTA
TGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCAT
CGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA
CTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCA
AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTGTGAAACTTCGAGGAGTCTCT
TTGTTGAGGACTTTTGAGTTCTCCCTTGAGGCTCCCACAGATACAATAAATATTTGAGAT
TGAACCCTGTCGAGTATCTGTGTAATCTTTTTTACCTGTGAGGTCTCGGAATCCGGGCCG
AGAACTTCGCAGTTGGCGCCCGAACAGGGACTTGATTGAGAGTGATTGAGGAAGTGAAGC
TAGAGCAATAGAAAGCTGTTAAGCAGAACTCCTGCTGACCTAAATAGGGAAGCAGTAGCA
GACGCTGCTAACAGTGAGTATCTCTAGTGAAGCGGACTCGAGCTCATAATCAAGTCATTG
TTTAAAGGCCCAGATAAATTACATCTGGTGACTCTTCGCGGACCTTCAAGCCAGGAGATT
CGCCGAGGGACAGTCAACAAGGTAGGAGAGATTCTACAGCAACATGGGGAATGGACAGGG
GCGAGATTGGAAAATGGCCATTAAGAGATGTAGTAATGTTGCTGTAGGAGTAGGGGGGAA
GAGTAAAAAATTTGGAGAAGGGAATTTCAGATGGGCCATTAGAATGGCTAATGTATCTAC
AGGACGAGAACCTGGTGATATACCAGAGACTTTAGATCAACTAAGGTTGGTTATTTGCGA
TTTACAAGAAAGAAGAGAAAAATTTGGATCTAGCAAAGAAATTGATATGGCAATTGTATA
CATGTGGAATCAGGATATATATGGGCACAAATAATTTCTCAAGAAACTGCTGACTGTACA
GTTAAAGCTGTCTTACAATTGTTGAGTGCTCATAATGTTACTGAATTACAAACAGATAAT
GGACCAAATTTTAAAAATCAAAAGATGGAAGGAGTAATTGTGACATTAAAAGTCTTTGCG
GTAGCAGGACTTTTAAATATGACGGCGCGCGTAATACGACTCACTATAGGGCGAATTGGG
TACCGCGATGTACGGGCCAGATATAGCATTTTGGGGTGACTCAGTAAATGGTAGATCTTG
CTACCAGTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCTAAGTGGTA
CTCTCCCAGAGACTGTCTGACTCACGCCACCCCCTCCACCTTGGACACAGGACGCTGTGG
TTTCTGAGCCAGGTACAATGACTCCTTTCGGTAAGTGCAGTGGAAGCTGTACACTGCCCA
GGCAAAGCGTCCGGGCAGCGTAGGCGGGCGACTCAGATCCCAGCCAGTGGACTTAGCCCC
TGTTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCCTCCCCCGT
TGCCCCTCTGGATCCACTGCTTAAATACGGACGAGGACAGGGCCCTGTCTCCTCAGCTTC
AGGCACCACCACTGACCTGGGACAGTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGG
TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCT
TCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAG
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCG
AGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA
AGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCT
ATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACA
TCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACG
GCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACC
CCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTC
TCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTG
CAGTCGACGGTACCGCGGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCC
ATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACC
TGAAACATAAAATGAATGCAATTGATCCAGCACAGTGGCGGCCGCTCGAGTGCGCCCTAA
TAGAAGCTAGTAATAGGATTAAAAGTAGAAGCTATGGAAAAATTTTTGTATACAGCTTTC
GCTATGCAAGAATTAGGATGTAATCAAAATCAATTTTTCTGCAAAATCCCTCCTGAGTTG
TGGACAAGGTATAATATGACTATAAATCAAACAATATGGAATCATGGAAATATAACTTTG
GGGGAATGGTATAACCAAACAAAAGATTTACAACAAAAGTTTTATGAAATAATAATGGAC

ATAGAACAAAATAATGTACAAGGGAAGAAAGGGATACAACAATTACAAAAGTGGGAAGAT
TGGGTAGGATGGATAGGAAATATTCCACAATATTTAAAGGGACTATTGGGAGGTATCTTG
GGAATAGGATTAGGAGTGTTATTATTGATTTTATGTTTACCTACATTGGTTGATTGTATA
AGAAATTGTATCCACAAGATACTAGGATACACAGTAATTGCAATGCCTGAAGTAGAAGGA
GAAGAAATACAACCACAAATGGAATTGAGGAGAAATGGTAGGCAATGTGGCATGTCTGAA
AAAGAGGAGGAATGATGAAGTATCTCAGACTTATTTTATAAGGGAGATACTGTGCTGAGT
TCTTCCCTTTGAGGAAGGTATGTCATATGAATCCATTTCGAATCAAATCAAACTAATAAA
GTATGTATTGTAAGGTAAAAGGAAAAGACAAAGAAGAAGAAGAAAGAAGAAAGCCTTCAT
CAAGCTTATCGATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCT
TAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGC
TATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCT
TTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGA
CGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGC
TTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGAC
AGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTT
TCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGT
CCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCC
TCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCC
GCCTGATCGATACCGTCGACCTCGAGGGGGGGCCCTCCTAAAAAAGAAAAAAGGGTGGAC
TGGGATGAGTATTGGAACCCTGAAGGCCTTAACCAGTGCTTTGTGAAACTTCGAGGAGTC
TCTTTGTTGAGGACTTTTGAGTTCTCCCTTGAGGCTCCCACAGATACAATAAATATTTGA
GATTGAACCCTGTCGAGTATCTGTGTAATCTTTTTTACCTGTGAGGTCTCGGAATCCGGG
CCGAGAACTTCGCACTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCT
GTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGT
AAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCC
GCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGG
AGAGGCGGTTTGCGTATTGGGCGCT

Sequence features:
Length 7585 bp
AmpR 1801-941
CMV enhancer/promoter 2871-3458
5′ LTR 3459-3611
FIV non-coding sequence 3612-3883
FIV gag remnant 1 3884-4135
(Gag ATG in red) 3884-3886
FIV pol remnant 4136-4296
FIV gag remnant 2 4297-4345
This it the theoretical bit…
human α1-antitrypsin promoter 4432-4826
remains of MluI site 4405
possible extra sequence 5′ to promoter 4406-4431
eGFP 4827-5624
possible extra sequence 3′ to GFP 5625-5724
remains of EcoRV site 5725-5727
RRE region 5767-6479
WPRE 6493-7085
(mutated X promoter) 6883-6903
(mutated WHx ATG) 6904-6906
3′ LTR 7141-7334
U5 7141-7165
site of deletion 7166-7181
R / U3 7182-7334

APPENDIX 3

Sequence for pSmart2FIX 2 vector

AGATCTTGAATAATAAAATGTGTGTTTGTCCGAAATACGCGTTTTGAGATTTCTGTCGCC
GACTAAATTCATGTCGCGCGATAGTGGTGTTTATCGCCGATAGAGATGGCGATATTGGAA
AAATTGATATTTGAAAATATGGCATATTGAAAATGTCGCCGATGTGAGTTTCTGTGTAAC
TGATATCGCCATTTTTCCAAAAGTGATTTTTGGGCATACGCGATATCTGGCGATAGCGCT
TATATCGTTTACGGGGGATGGCGATAGACGACTTTGGTGACTTGGGCGATTCTGTGTGTC
GCAAATATCGCAGTTTCGATATAGGTGACAGACGATATGAGGCTATATCGCCGATAGAGG
CGACATCAAGCTGGCACATGGCCAATGCATATCGATCTATACATTGAATCAATATTGGCC
ATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCA
TACGTTGTATCCATATCGTAATATGTACATTTATATTGGCTCATGTCCAACATTACCGCC
ATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCA
TAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACC
GCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAAT
AGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGT
ACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCC
CGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTA
CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGG
ATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTT
GTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTGCGATCGCCCGCC
CCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGT
TTAGTGAACCGGGCACTCAGATTCTGCGGTCTGAGTCCCTTCTCTGCTGGGCTGAAAAGG
CCTTTGTAATAAATATAATTCTCTACTCAGTCCCTGTCTCTAGTTTGTCTGTTCGAGATC
CTACAGTTGGCGCCCGAACAGGGACCTGAGAGGGGCGCAGACCCTACCTGTTGAACCTGG
CTGATCGTAGGATCCCCGGGACAGCAGAGGAGAACTTACAGAAGTCTTCTGGAGGTGTTC
CTGGCCAGAACACAGGAGGACAGGTAAGATTGGGAGACCCTTTGACATTGGAGCAAGGCG
CTCAAGAAGTTAGAGAAGGTGACGGTACAAGGGTCTCAGAAATTAACTACTGGTAACTGT
AATTGGGCGCTAAGTCTAGTAGACTTATTTCATGATACCAACTTTGTAAAAGAAAAGGAC
TGGCAGCTGAGGGATGTCATTCCATTGCTGGAAGATGTAACTCAGACGCTGTCAGGACAA
GAAAGAGAGGCCTTTGAAAGAACATGGTGGGCAATTTCTGCTGTAAAGATGGGCCTCCAG
ATTAATAATGTAGTAGATGGAAAGGCATCATTCCAGCTCCTAAGAGCGAAATATGAAAAG
AAGACTGCTAATAAAAAGCAGTCTGAGCCCTCTGAAGAATATCTCTAGAGTCGACGCTCT
CATTACTTGTAACAAAGGGAGGGAAAGTATGGGAGGACAGACACCATGGGAAGTATTTAT
CACTAATCAAGCACAAGTAATACATGAGAAACTTTTACTACAGCAAGCACAATCCTCCAA
AAAATTTTGTTTTTACAAAATCCCTGGTGAACATGGTCGACTCTAGAACTAGTGGATCCC
CCGGGCTGCAGGAGTGGGGAGGCACGATGGCCGCTTTGGTCGAGGCGGATCCGGCCATTA
GCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACG
TTGTATCCATATCATAATATGTACATTTATATTGGCTCATGTCCAACATTACCGCCATGT
TGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGC
CCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCC
AACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGG
ACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACAT
CAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCC
TGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAG
CGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT
TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAA
ATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT

**CAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGA
TCCAGCCTCCGCGGCCCCACTAGAACTAGTGGATCCGTCGACCATGCAGCGCGTGAACAT
GATCATGGCAGAATCACCAGGCCTCATCACCATCTGCCTTTTAGGATATCTACTCAGTGC
TGAATGTACAGTTTTTCTTGATCATGAAAACGCCAACAAAATTCTGAATCGGCCAAAGAG
GTATAATTCAGGTAAATTGGAAGAGTTTGTTCAAGGGAACCTTGAGAGAGAATGTATGGA
AGAAAAGTGTAGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGAAAGAACAACTGA
ATTTTGGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGG
CAGTTGCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAA
GAACTGTGAATTAGATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAA
AAATAGTGCTGATAACAAGGTGGTTTGCTCCTGTACTGAGGGATATCGACTTGCAGAAAA
CCAGAAGTCCTGTGAACCAGCAGTGCCATTTCCATGTGGAAGAGTTTCTGTTTCACAAAC
TTCTAAGCTCACCCGTGCTGAGACTGTTTTTCCTGATGTGGACTATGTAAATTCTACTGA
AGCTGAAACCATTTTGGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCACTCG
GGTTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCCTTGGCAGGTTGTTTTGAATGG
TAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAACTGCTGC
CCACTGTGTTGAAACTGGTGTTAAAATTACAGTTGTCGCAGGTGAACATAATATTGAGGA
GACAGAACATACAGAGCAAAAGCGAAATGTGATTCGAATTATTCCTCACCACAACTACAA
TGCAGCTATTAATAAGTACAACCATGACATTGCCCTTCTGGAACTGGACGAACCCTTAGT
GCTAAACAGCTACGTTACACCTATTTGCATTGCTGACAAGGAATACACGAACATCTTCCT
CAAATTTGGATCTGGCTATGTAAGTGGCTGGGGAAGAGTCTTCCACAAAGGGAGATCAGC
TTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACCGAGCCACATGTCTTCGATCTAC
AAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAGATTC
ATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGGAAGGGACCAGTTTCTTAAC
TGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATACCAA
GGTATCCCGGTATGTCAACTGGATTAAGGAAAAAACAAAGCTCACTTAATGAAAGATGGA
TTTCCAAGGTTAATTCATTGGAATTGAAAATTAACAGGGCCTCTCACTAACTAATCACTT
TCCCATCTTTTGTTAGATTTGAATATATACATTCTATGATCATTGCTTTTTCTCTTTACA
GGGGAGAATTTCATATTTTACCTGAGCAAATTGATTAGAAAATGGAACCACTAGAGGAAT
ATAATGTGTTAGGAAATTACAGTCATTTCTAAGGGCCCAGCCCTTGACAAAATTGTGAAG
TTAAATTCTCCACTCTGTCCATCAGATACTATGGTTCTCCACTATGGCAACTAACTCACT
CAATTTTCCCTCCTTAGCAGCATTCCATCTTCCCGATCTTCTTTGCTTCTCCAACCAAAA
CATCAATGTTTATTAGTTCTGTATACAGTACAGGATCCCCCGGGCTGCAGGAATTCGATA
TCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGACCTACTAGGGTGCTGTGGAA
GGGTGATGGTGCAGTAGTAGTTAATGATGAAGGAAAGGGAATAATTGCTGTACCATTAAC
CAGGACTAAGTTACTAATAAAACCAAATTGAGTATTGTTGCAGGAAGCAAGACCCAACTA
CCATTGTCAGCTGTGTTTCCTGACCTCAATATTTGTTATAAGGTTTGATATGAATCCCAG
GGGGAATCTGAACCCCTATTACCCAACAGTCAGAAAAATCTAAGTGTGAGGAGAACACAA
TGTTTCAACCTTATTGTTATAATAATGACAGTAAGAACAGCATGGCAGAATCGAAGGAAG
CAAGAGACCAAGAATGAACCTGAAAGAAGAATCTAAAGAAGAAAAAAGAAGTGGAAAATA
GGTATGTTTCTGTTATGCTTAGCAGGAACTACTGGAGGAATACTTTGGTGGTATGAAGGA
CTCCCACAGCAACATTATATAGGGTTGGTGGCGATAGGGGGAAGATTAAACGGATCTGGC
CAATCAAATGCTATAGAATGCTGGGGTTCCTTCCCGGGGTGTAGACCATTTCAAAATTAC
TTCAGTTATGAGACCAATAGAAGGTGACCAGTGGTGCAGGGTCCTCCGGCAGTCGTTACC
TGAAGAAAAAATTCCATCACAAACATGCATCGCGAGAAGACACCTGGGACCAGGCCCAAC
ACAACATACACCTAGCAGGCGTGACCGGTGGATCAGGGGACAAATACTACAAGCAGAAGT
ACTCCAGGAACGACTGGAATGGAGAATCAGAGGAGTACAACAGGCGGCCAAAGAGCTGGG
TGAAGTCAATCGAGGCATTTGGAGAGAGCTATATTTCCGAGAAGACCAAAGGGGAGATTT
CTCAGCCTGGGGCGGCTATCAACGAGCACAAGAACGGCTCTGGGGGGAACAATCCTCACC
AAGGGTCCTTAGACCTGGAGATTCGAAGCGAAGGAGGAAACATTTATGACTGTTGCATTA
AAGCCCAAGAAGGAACTCTCGCTATCCCTTGCTGTGGATTTCCCTTATGGCTATTTTGGG
GGCTAGTAATTATAGTAGGACGCATAGCAGGCTATGGATTACGTGGACTCGCTGTTATAA
TAAGGATTTGTATTAGAGGCTTAAATTTGATATTTGAAATAATCAGAAAAATGCTTGATT
ATATTAGCTTCGAATCACTAGTGAATTCGCGGCCGCAATCAACCTCTGGATTACAAAAT
TTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGC**

TGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTT
GTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGG
CGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTG
TCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGC
CGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGT
GTTGTCGGGGAAGCTGACGTCCTTTCCATGGCTGCTCGCCTGTGTTGCCACCTGGATTCT
GCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCG
CGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCG
GATCTCCCTTTGGGCCGCCTCCCCGCCTGATCGATGTCGAGAATCACTAGTGAATTCGCG
GCCGCTCGACAGCTTATCGATACCGTCGAATTGGAAGAGCTTTAAATCCTGGCACATCTC
ATGTATCAATGCCTCAGTATGTTTAGAAAAACAAGGGGGGAACTGTGGGGTTTTTATGAG
GGGTTTTATAAATGATTATAAGAGTAAAAAGAAAGTTGCTGATGCTCTCATAACCTTGTA
TAACCCAAAGGACTAGCTCATGTTGCTAGGCAACAATTGGGCACTCAGATTCTGCGGTCT
GAGTCCCTTCTCTGCTGGGCTGAAAAGGCCTTTGTAATAAATATAATTCTCTACTCAGTC
CCTGTCTCTAGTTTGTCTGTTCGAGATCCTACAGAGCTCATGCCTTGGCGTAATCATGGT
CATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCG
GGAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGT
TGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCG
GCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTG
ACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAA
TACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGC
AAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCC
CTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTAT
AAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGC
CGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCT
CACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACG
AACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACC
CGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGA
GGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA
GAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTA
GCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGC
AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTG
ACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGA
TCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATG
AGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCT
GTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGG
AGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTC
CAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAA
CTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGC
CAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGT
CGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCC
CCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGT
TGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGC
CATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGT
GTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATA
GCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGA
TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAG
CATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA
AAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATT
ATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGA
AAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTAAATTGTAAG
CGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCA
ATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAG
TGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGG
GCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTT

TTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAG AGCTTGACGGGGAAAGCCAACCTGGCTTATCGAAATTAATACGACTCACTATAGGGAGAC CGGC

Sequence features: Length 9434 bp
CMV enhancer 543-1150
5' LTR 1151-1264
R 1151-1225
U5 1226-1265
tat exon 1 1308-1403
major splice donor CAG*GTAAGA 1403
gag remnant 1409-1783
(ATGs mutated to ATTGs 1409 & 1427)
pol remnant 1796-1955
(includes cPPT 1811-1824)
CMV promoter 2030-2838
FIX coding sequence 2864-4251
FIX 3' UTR 4252-4657
pol remnant 4719-4831
major splice acceptor TTGTTGCAG*GAA 4844
tat exon 2 remnant 4844-4901
s2 remnant 4914-5057
rev exon 1 remnant 5022-5110
rev exon 2 remnant 5318-5699
env remnants 4914-5892
6579-6635
WPRE 5927-6519
(unmutated X promoter 6317-6337)
(unmutated X start codon 6338-6349)
3′ PPT 6637-6650
3' LTR 6655-6883
U3 6654-6768
R 6769-6843
U5 6844-6883
AmpR 8930-8070

pSmart2FIX 2 sequenced 2005
Predicted amino acid sequence of Factor IX:
MQRVNMIMAE SPGLITICLL GYLLSAECTV FLDHENANKI LNRPKRYNSG
KLEEFVQGNL ERECMEEKCS FEEAREVFEN TERTTEFWKQ YVDGDQCESN
PCLNGGSCKD DINSYECWCP FGFEGKNCEL DVTCNIKNGR CEQFCKNSAD
NKVVCSCTEG YRLAENQKSC EPAVPFPCGR VSVSQTSKLT RAETVFPDVD
YVNSTEAETI LDNITQSTQS FNDFTRVVGG EDAKPGQFPW QVVLNGKVDA
FCGGSIVNEK WIVTAAHCVE TGVKITVVAG EHNIEETEHT EQKRNVIRII
PHHNYNAAIN KYNHDIALLE LDEPLENTIVIRUSESLNSY VTPICIADKE
YTNIFLKFGS
GYVSGWGRVF HKGRSALENTIVIRUSESLQ YLRVPLENTIVIRUSESDRA
TCLRSTKFTI YNNMFCAGFH
EGGRDSCQGD SGGPHVTEVE GTSFLTGIIS WGEECAMKGK YGIYTKVSRY
VNWIKEKTKL T•
(461 residues)

APPENDIX 4

Observations in MF-1 mice treated with SMART2 vector

Vector	Mouse tissue samples	Gender	Phenotype at death
-	Uninfected	Male	*
-	Cat	Male	Well appearance. Mouse tissues normal (age at death = 715).
-	Ching	Female	Large amounts of adipose tissue, large blood clot on the side of neck otherwise tissues appeared normal (age at death = 819).
SMART2Z	Angus	Male	Medium sized liver tumour palpable, very small lung tumour. Bleb on liver near gall bladder
	Agnus	Female	Tumour palpable, very small lung tumour. Bleb on liver near gall bladder
	Archie	Male	Mouse very sick, Single small liver tumour
	Adame	Female	Solid mass attached to gut and four liver tumours. Lungs appeared wasted.
SMART2ZWΔ	Gina	Male	Well appearance, liver tumour by palpation
	Gerty	Female	Very small and ill looking (half weight of other mice), no tumour found
	Golliath	Male	Very sick, large kidneys, two liver tumours and multiple liver lumps, large spleen with blemishes
	Gladys	Female	Well appearance but tumour palpable
	Gerry	Male	Healthy but a big bulgy, 8 liver tumours found
	Gonzo	Male	Scabby, sore patches on face and ears. Hunch swollen abdomen fir on end. Six bone tumours on hind limbs. No abnormal liver or lungs.
SMART2 hFIX	Jonas	Male	Well appearance but tumour palpable
SMART2lacZ LCMV	Monika	Male	Well appearance but large tumour palpable
SMART2lacZ rabies	Rachel	Female	Well appearance, no tumour palpable
SMARTZ Ebola	Mike	Male	Ungroomed, huge tumour palpable

* Is used to represent mice for which the phenotype was not given for this study.

APPENDIX 5

Observations in MF-1 mice treated with pLIONhAATGFP vector

Vector	Mouse tissue samples	Gender	Phenotype at death
pLIONhAATGFP	Flora	Female	Scabby back of neck and ears and conjunctivitis in one eye. All tissue normal except for an ovarian cyst
	Frank	Male	Healthy looking but tumour visibly large and easily palpable
	Fanny	Female	There were two tumours. Tumour 1 glowed intense green with two extremely green foci, and tumour 2 glowed much less, but had foci or seams of more intense glowing than the surrounding liver.
	Fiona	Female`	Healthy looking but two floppy green tumours inside on liver

APPENDIX 6

Observations in MF-1 mice treated with HR'SIN-cPPT-S-FIX-W vector

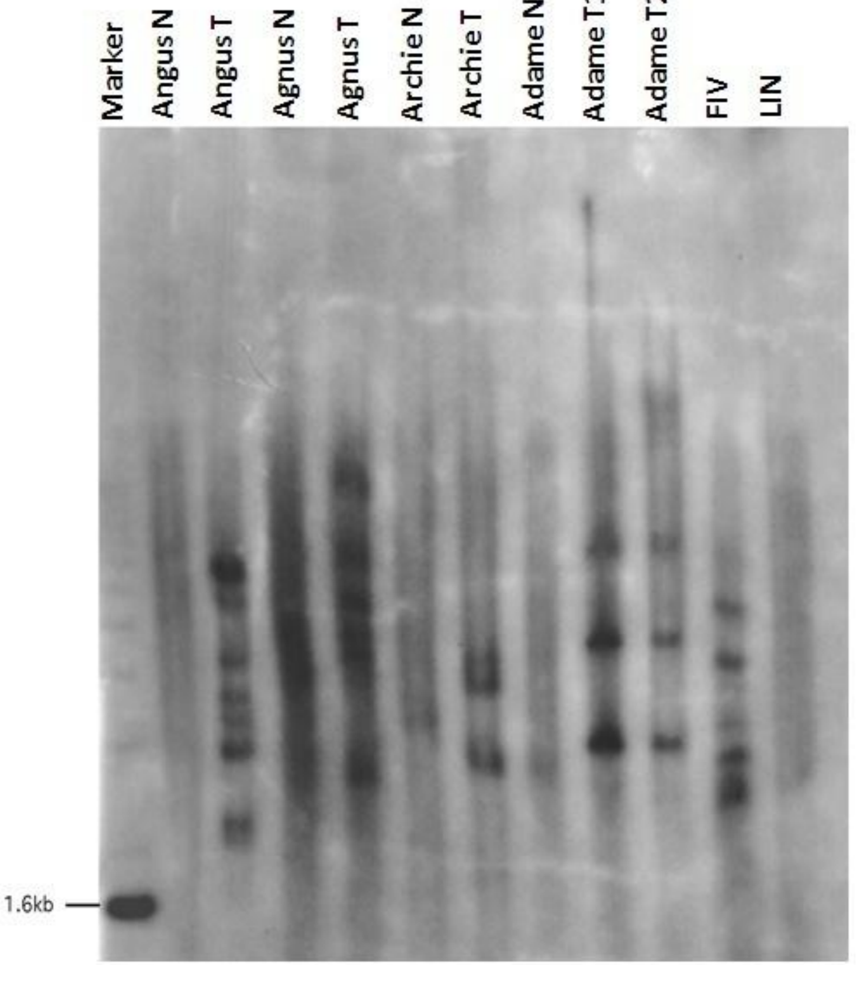
Vector	Mouse tissue samples	Gender	Phenotype at death*
HR'SIN-cPPT-S-FIX-W	Topsy	Female	Neck mass was soft and necrotic, One dark ovary enlarged and one small also dark very bloody, Very enlarged spleen,
	Tilley	Female	Mouse tissues normal. Lung found with large tumour
	Kostas	Male	Scabby mouse around shoulder blades. Otherwise healthy. Liver normal.
	Bruce	Female	*
	Pierre	Male	*

* Is used to represent mice for which the phenotype was not given for this study.

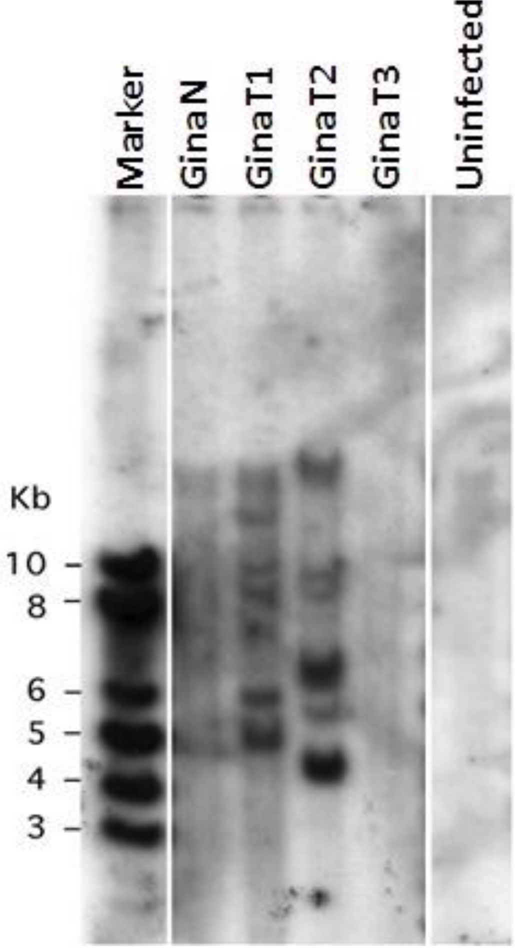
APPENDIX 7

Southern analysis was also carried out prior to this thesis to determine the clonality of tumours. Southern analysis was performed on normal hepatic tumour tissue that was isolated from SMART2Z, SMART2ZWΔ and pLIONhAATGFP treated MF-1 mice. This study showed that clonally derived liver tumours develop in mice following lentivirus delivery into fetal mice. VCN's ranging from 1-10 was shown with varying intensities of vector signal following hybrisdisation of a radioactive WPRE probe to equal amounts of tumour DNA. The results produced by normal samples showed a smear. This confirms the heterogenous nature of the normal samples which contain numerous provirus insertions per sample.

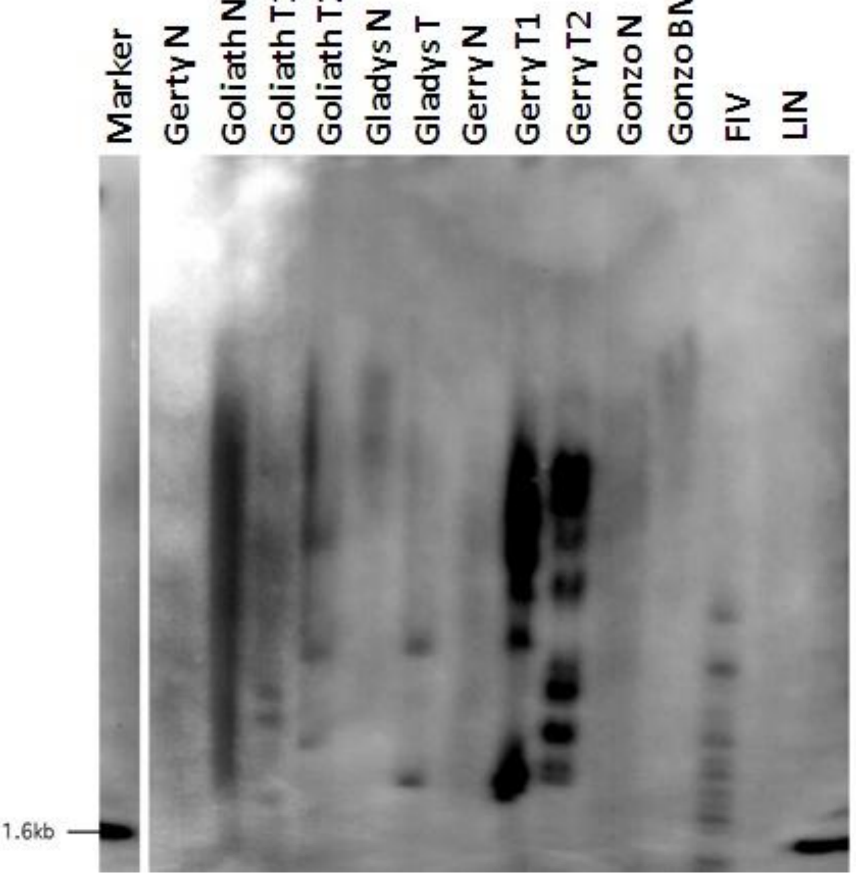
SMART2Z



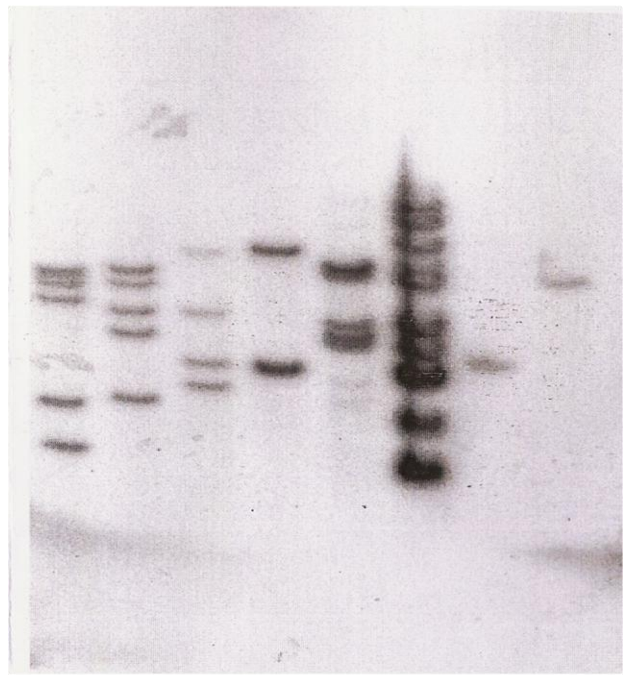
SMART2ZWΔ



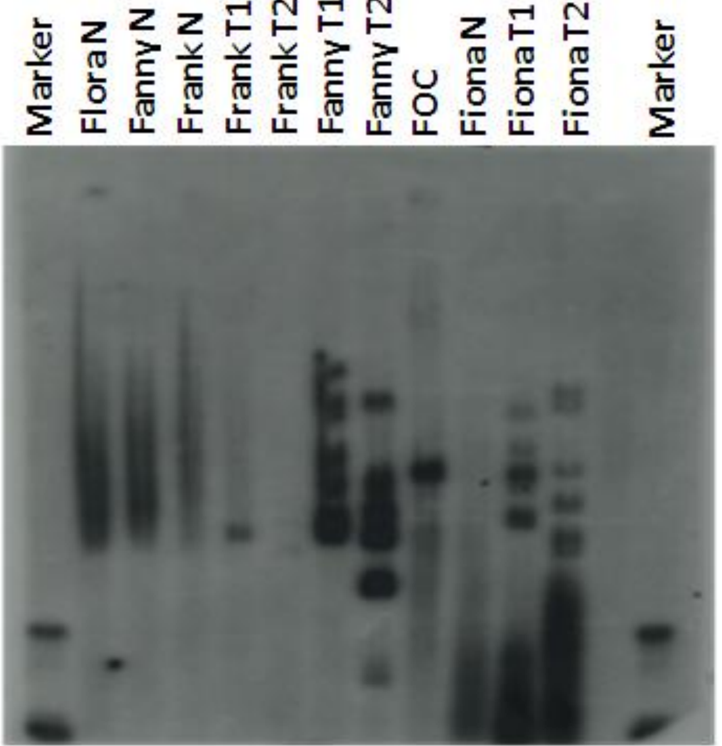
SMART2ZWΔ



SMART2ZWΔ



pLIONhAATGFP



The restriction enzyme *Hind III* was used to geneate DNA fragments . Integrated proviruses were detected and identified by hibridisation with a α-^{32}P-CTP labelled WPRE probe (625bp). Vector copy numbers were estimated according to the visibility of bands on autoradiographs following hybridisation. Intense fragments correlated to multiple bands. No signal was obtained from control non-infected hepatic issue samples.

APPENDIX 8

FIV integration sites sequences

Flora Ovarian Cyst 1,166bp

```
CCAAACTCTATACGACTCACTATAGGGAAAGCTCGGTACCACGCATGCTGCAGACGCGTT
ACGTATCGGATCCAGAATTCGTGATTCTCAAAAGTCCTCAACAAAGGAGTACACGTGGCT
CCAGCTTCATAGGTAGCAGAGGGTGGCTTTGTCATGCATCAGTGGAAGGAGAGGTCCTTG
GTCCTATGAAGGCTTGATAGATGGCCCATTGTAGGGTAATTGAGGGCAGGGAGGTGGGAG
TTGTTAGGTGGGTAGAGAAACACACTCATAGAAGCATGGAGAGGGAGGTTGAGATAGGGG
TTTTCTGGGAGGGAGGAAAACCAGGAAAGGGGATACCATTTGAAATGTAAATAAGGAAAA
TATCCTTTAAAAAATAAAACCTAACTGCTGTGCCACTGAATTCAGATCAATCTGAATTCG
TCGACAAGCTTCTCGAGCCTAGGCTAGCTCTAGACCACACGTGTGGGGGCCCGAGCTCGC
GGCCGCTGTATTCTATAGTGTCACCTAAATGGCCGCACAATTCACTGGCCGTCGTTTTAC
AACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCC
CTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGC
GCAGCCTGAATGGCGAATGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAA
TTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAA
ATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACT
ATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCC
ACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAA
ATCGGAACCCTAAAGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGACGTGGCG
AGAAAGGAGGGAGAAAGCGAAGGAGCGGTCGCTTAGGCGCTGGCA
```

10 – 85 = pDrive vector
98 – 379 = Chr 13 (*Prl4a1* or *Prl2c3*)
380 – 408 = LC3 linker
410 – 1065 = pDrive vector

Flora Ovarian Cyst 1,281bp

```
CCAGCCCTATACGACTCACTATAGGGAAAGCTCGGTACCACGCATGCTGCAGACGCGTTA
CGTATCGGATCCAGAATTCGTGATGATCTGAATTCAGTGGCACAGCAGTTAGGGATCATG
AAGGCCATCTGCAGCTATCTTTGTTGAGGACTTTTGAGAATCTGAATTCGTCGACAAGCT
TCTCGAGCCTAGGCTAGCTCTAGACCACACGTGTGGGGGCCCGAGCTCGCGGCCGCTGTA
TTCTATAGTGTCACCTAAATGGCCGCACAATTCACTGGCCGTCGTTTTACAACGTCGTGA
CTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAG
CTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAA
TGGCGAATGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAA
ATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAA
TAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAAC
GTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAA
CCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCT
AAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAA
GGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGC
GTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTT
CGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTAT
CCGCTCATGAGACATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAGAGTATGAG
TATTCAACATTTCCGTGTCGCCCTTATCCCTTTTTTGCGGCATTTTGCTTCCTGTTTTGC
TCACCCAGAACGCTGGTGAAGTAAAGATGCTGAGATCAGTTGGGTGCCACGAGTGCTTAC
ATCGAACTGATCTTAACAGCG
```

9 – 84 = pDrive vector
85 – 113 = LC1 linker
118 – 138 = Chr14 (*Sacs/Tnfrsf19*)
139 – 158 = FIV
160 – 1161 = pDrive vector

Fiona T2 **1,177bp**

GCGCTCTATACGACTCACTATAGGGAAAGCTCGGTACCACGCATGCTGCAGACGCGTTAC
GTATCGGATCCAGAATTCGTGATTGATCTGAATTCAGTGGCACAGCAGTTAGGGATCATA
AAGATCTAGCTTTAAATACCTCACAAGAATTTTAAATAAATATGGTCCTCCTCTCCCATC
AGTATCCTACACAAAATGCCACCTGCTTAAATACTTCCTAAAAAAGAGCTGGGGAGAGAA
CTTCACTGGGGAAATGGGATGAGTATTGGAACCCTGAAGGCCCGATTGATCAGCTCCTAG
GTAACCAGTGCTTTGTGAAACTTCGAGGAGTCTCTTTGTTGAGGACTTTTGAGAATCTGA
ATTCGTCGACAAGCTTCTCGAGCCTAGGCTAGCTCTAGACCACACGTGTGGGGGCCCGAG
CTCGCGGCCGCTGTATTCTATAGTGTCACCTAAATGGCCGCACAATTCACTGGCCGTCGT
TTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACA
TCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACA
GTTGCGCAGCCTGAATGGCGAATGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCG
TTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCT
TATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGT
CCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGAT
GGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGC
ACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAA
CGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGCGCTGGCAAGTGTA
GCGGTCACGCTGCGCGTACCACCACACCCGCCGCGCTTATGCGCCGCTAGAGGCGCGTCA
GTGCACTTTTCGGGGAAATGTG

8 – 83 = pDrive vector
85 – 113 = LC1 linker
116 – 254 = Chr 5 (*Pscd3*)
255 – 353 = FIV
355 – 1102 = pDrive vector

Fanny T1 **1,159bp**

ACGTCTTATACGACTCACTATAGGGAAGCTCGGTACCACGCATGCTGCAGACGCGTTACG
TATCGGATCCAGAATTCGTGATTCTCAAAAGTCCTCAACAAAGAGACTCCTCGAAGTTTC
ACAAAGCACTGGTTACCTAGGAGCTGATCAATCAGGCCTTCAGGGTTCCAATACTCATCC
CAGTAACAAGTTAAACAATGACATCAAAAAGACTTGAAGATAAGGGACATTACTACCATA
CTCCAAAATAAAAAGAATGGTTCAAAATGCCTTTCCCCCACTCTCCTCTGTCCACTGTAC
TGTCAGTTCACCTTTAGGATTTGTCTTGGAGCTATTGATCCCTAACTGCTGTGCCACTGA
ATTCAGATCAATCTGAATTCGTCGACAAGCTTCTCGAGCCTAGGCTAGCTCTAGACCACA
CGTGTGGGGGCCCGAGCTCGCGGCCGCTGTATTCTATAGTGTCACCTAAATGGCCGCACA
ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTA
ATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCG
ATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGAAATTGTAAGCGTTAATAT
TTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGA
AATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCC
AGTTTGGAACAAGAGTCCACTATTAAGAACGTGGACTCCAACGTCAAAGGGCGAAAACCG
TCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTGGGGTCGAG
GTGCCGTAAAGCACTAAATCGGACCCTAAAGGGAGCCCCGATTTAGAGCTTGACGGGGAA
GCCGGCGACGTGGCGAGAAGGAGGGAGAAGCGAAAGGGCGGGCGCTAGGCGCTGCAAGTG
TAGCGTCACGCTGCGCGTAACCACACACCGCGCGCTTATGCGCCGCTACAGGCGCGTCAG

8 – 82 = pDrive vector
84 – 182 = FIV
183 – 336 = Chr 18 (*Cx39/similar to cyclin fold protein1*)
341 – 369 = LC3 linker
371 – 1080 = pDrive vector

Fanny T1 **684bp**

**GGCGGCCCTATACGACTTCCTTATAGGGAAAGCTCGGTACCACGCATGCTGCAGACGCGT
TACGTATCGGATCCAGAATTCGTGATTGATCTGAATTCAGTGGCACAGCAGTTAGGAATT
ATATTAAACCTAGTGATGAGGATAAGAGAAAGAAACTTGCTAGGCAGTTAAGGGAAAGAA
ACGTTGCAAACAAGCCCTTCTGCCGTGGCCAAAAGTATCCTTGGCTCCAATGACAGTTCT
GGTCTGGCCAGAGGGCAGAACATACTGGTTTGTGTTCTTAGATGTTTACTTGTCTTTTTT
CTTTAGTACAGTAATCTAATGAGCATAGCTGCCTGGAAGCTAGACATACTCTAATACTTC
ATTGGTCCTTGGAGGTGTTTGGGATGAGTATTGGAACCCTGAAGGCCTGATTGATCAGCT
CCTAGGTAACCAGTGCTTTGTGAAACTTCGAGGAGTCTCTTTGTTGAGGACTTTTGAGAA
TCTGAATTCGTCGACAAGCTTCTCGAGCCTAGGCTAGCTCTAGACCACACGTGTGGGGGC
CCGAGCTCGCGGCCGCTGTATTCTATAGTGTCACCTAAATGGCCGCACAATTCACTGGCC
GTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCA
GCACATCCCCCTTTCGCCAGCTGG**

22 – 86 = pDrive vector
88 – 116 = LC1 linker
117 – 380 = Chr 19 (*Pdcd1lg2*)
381 – 478 = FIV
480 – 684 = pDrive vector

Frank T1 **853bp**

**ACTAAATTACGGGGGAAAAGCTCGGGTACCACGCATGCTAGCAGACGCGTTACGTATCGG
ATCCAGAATTCGTGATCTCAAAAGTCCTCAACAAAGAGACTCCTCGAAGTTTCACAAAGC
ACTGGTTACCTAGGAGCTGATCAATCAGGCCTTCAGGGTTCCAATACTCATCCCACTTGG
GAACACACTGTTGGCTACACAATCCTAGTTTTGATCTCTGTCTTAGGATTCCCCTCATTC
CATCTAATAGCAGCATCTCCCCTCTCCCAGACACTCTGACATAGTCTTATTCCATAGTCA
CTAGTCACTCACTTCCATGAACTCTTAAGTGCCTCCACTTCTCCCTGTGACATCTGTACC
CTCTTAGCTCTATAAACATCTTGCCTTGTCTGTATTGTCACAAATCTTAAACTCTCATTA
CCCTCACCATTTTGTTCCCAAAGAATCATCCAAGTTGCCTTTCTCAAATACAAATCATGT
AAAATCTTGCTCTTAGTCCTTTGTGCAGAATCTAAACTCCCAATTCCTAACTGCTGTGCC
ACTGAATTCAGATCAATCTGAATTCGTCGACAAGCTTCTCGAGCCTAGGCTAGCTCTAGA
CCACACGTGTGGGGGCCCGAGCTCGCGGCCGCTGTATTCTATAGTGTCACCTAAATGGCC
GCACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCA
ACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTATAGCGAAAGAGGCCCG
CACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAAATGGCGAATGGAATTGTAAGCGTT
ATATTTTGTTAAA**

25 – 76 = pDrive vector
176 – 262 = Chr12 (*Nek9*)
263 – 521 = Chr 3 (*Pglyrp4*)
522 – 554 = LC3 linker
556 – 853 = pDrive vector

Frank T1 **1,063bp**

```
AACTTAATAAGGGGGAAAAGCTTCGGTACCACGCATGCTGCAGACGCGTTACGTATCGGA
TCCAGAATTCGTGATTCTCAAAAGTCCTCAACAAAGAGACTCCTCGAAGTTTCACAAAGC
ACTGGTTACCTAGGAGCTGATCAATCAGGCCTTCAGGGTTCCAATACTCATCCCACTTGG
GAACACACTGTTGGCTACACAATCCTAGTTTTGATCTCTGTCTTAGGATTCCCCTCATTC
CATCTAATAGCAGCATCTCCCCTCTTCCTGACATGTACTGACAGTCCTACTATGAAATCA
GCTAACGTGGAACTGTTGCCACATCTCTGAGTAGTCAGATACCTCGGGCCCCTGCCTAGT
AGAGATAATGCTATATTTTTGTTTTAATTCCTAACTGCTGTGCCACTGAATTCAGATCAA
TCTGAATTCGTCGACAAGCTTCTCGAGCCTAGGCTAGCTCTAGACCACACGTGTGGGGGC
CCGAGCTCGCGGCCGCTGTATTCTATAGTGTCACCTAAATGGCCGCACAATTCACTGGCC
GTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCA
GCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCC
CAACAGTTGCGCAGCCTGAATGGCGAATGGAAATTGTAAGCGTTAATATTTTGTTAAAAT
TCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAA
TCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACA
AGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGG
GCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTA
AAGCACTAATTCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGG
```

23 – 75 = pDrive vector
77 – 175 = FIV
176 – 385 = Chr 12 (*Nek9*)
386 – 418 = LC3 linker
420 – 1011 = pDrive vector

Frank T2 **1,268bp**

```
ACGTCTTATCGACTCACTATAGGGAAGCTCGGTACCACGCATGCTGCAGACGCGTTACGT
ATCGGATCCAGAATTCGTGATTGATCTGAATTCAGTGGCACAGCAGTTAGGAGATAGCTG
AGACTCAGGTGGACTGGGATGAGTATTGGAACCCTGAAGGCCTGATTGATCAGCTCCTAG
GTAACCAGTGCTTTGTGAAACTTCGAGGAGTCTCTTTGTTGAGGACTTTTGAGAATCTGA
ATTCGTCGACAAGCTTCTCGAGCCTAGGCTAGCTCTAGACCACACGTGTGGGGGCCCGAG
CTCGCGGCCGCTGTATTCTATAGTGTCACCTAAATGGCCGCACAATTCACTGGCCGTCGT
TTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACA
TCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACA
GTTGCGCAGCCTGAATGGCGAATGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCG
TTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCT
TATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGT
CCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGAT
GGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCA
CTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAAC
GTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTA
GCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCG
TCAGTGGCACTTTTCGGGGATATGTGCGCGGACCGCTATTTGTTATTTTTCGGACTACAT
TCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCA
```

10 – 81 = pDrive vector
83 – 111 = LC1 linker
112 – 132 = Chr16 (*Coro7*)
133 – 233 = FIV
234 – 1067 = pDrive vector

The BLAST sequence results were used to annotate the mouse genomic DNA flanking the provirus integration site. Purple, red and green were used to represent the pDRIVE cloning plasmid, mouse genomic DNA and linker cassette sequence (used to ligate restriction products) respectively. Mouse genomic sequences represented in LAM PCR products included – a) *Prl4a1/ Prl2c3* b) *Sacs/Tnfrs19* c) *Pscd3*, d) *CX39/similar to cyclin fold protein1* e) *Pdcd1lg2* f) *Pglyrp4* g) *Nek9* h) *Coro7*.

APPENDIX 9

454 pyrosequencing results: EIAV integration sites with high sequence counts

Pah

**GATCTGAATTCAGTGGCACAGCAGTTAGGTTTCCAGAGCCTCCTGCTCCCTCCTAG
ATAGAATCTTTCAGTTTGGAAAAACTGCCATTCAGCAGCCTGAATTTGTGGCACTGA
GGAATAGAACCAAAGTGAAGAGTCTATTGTTTAGAACCAAACTCAGTTGATCCACTT
GTTATCGAGATGGGCAAAAGGAAGCTTTTAACCTCTTGGCTTAGTTTCCTTAGTTTC
TGTGGGGTTTTTATGAGGGGTTTTATAAATGATTATAAGAGTAAGAAGAAAGTTGCT
GATGCTCTCATAACCTTGTATAAC**

Acvr2a

**GATCTGAATTCAGTGGCACAGCAGTTAGGTGACAATTGAAGTTTAGAGATTAGGCA
CACCCAAGTCAGAAGGCAGATGAGACAGAGCATGAGCTGAAACATTGCATTTGGTT
TTTGATTTGTTTTGTTTTGTTTTATTGCTGGTGTCATTGTTTGGTTGGGTCTTTATTT
TGATTTTCTTTGTTTTGTTCACTGAGACAAGATCTCACCATGCAACTGTGGGGTTTT
TATGAGGGGTTTTATAAATGATTATAAGAGTAAGAAGAAAGTTGCTGATGCTCTCAT
AACCTTGTATAAC**

Setd5

**GATCTGAATTCAGTGGCACAGCAGTTAGGAATTCATGTTCTGGTGCTTCTAGTTAAG
TGTGTGGGGTTTTTATGAGGGGTTTTATAAATGATTATAAGAGTAAGAAGAAAGTT
GCTGATGCTCTCATAACCTTGTATAAC**

Mrpl1

**GATCTGAATTCAGTGGCACAGCAGTTAGGAATTTGTTTATGGTTATTGCTGAAGTAT
GTGCACAAAAATATTTTTTAAATATAAAAACTAATTTGAGCACTTTGGAAAGCCATG
ATACAGATGTAGCCAAGCTACGTCACCACCGTTGCTGCCGGATTTTGGATAAGAAA
ACAACCGAGGTCATAACTGATAGAGCACTTGCCTAATATTTGAAAGGCTCCTGGGG
TTTTTATGAGGGGTTTTATAAATGATTATAAGAGTAAGAAGAAAGTTGCTGATGCTC
TCATAACCTTGTATAAC**

Uvrag

**GATCTGAATTCAGTGGCACAGCAGTTAGGTTATGTTGTGGAATGTGACTTTGATCTA
CTTGGCAGGAATTTTATTTTACAAAACTTACTATTAAAAAAGACAGGTTTTCCATCC
TTTAATTATGCTGTGTTCACTGAACTTTGTTGGGGTTTTTATGAGGGGTTTTATAAA
TGATTATAAGAGTAAGAAGAAAGTTGCTGATGCTCTCATAACCTTGTATAAC**

Park7

**GATCTGAATTCAGTGGCACAGCAGTTAGGAATTGAACTCAGGACACCTGGAAGAGC
AGTCAGTGCTCTTAAGTGCTCTTAACTACTGAGTCATCTCTCCAACCCCAAAGGTCC
CTTACGTTTGGCCATCTTGGGGTTTTTATGAGGGGTTTTATAAATGATTATAAGAGT
AAGAAGAAAGTTGCTGATGCTCTCATAACCTTGTATAAC**

Integration sites in tumour derived EIAV infected mice. 454 sequencing was conducted by Manfred Schmidt's research group (Mannheim, Germany) on LAM PCR & non-restrictive PCR amplificons using the GS FLX system according to the manufactures instructions (Roche, Applied Science, USA). All sequences were conducted using Blas2seq & the Smith-Waterman algorithm. The raw sequences reads were aligned to the mouse genome using USL BLAT genome browser http://genome.ucsc.edu & the BLAST database.

APPENDIX 10 Genomic distribution of EIAV integrations with high sequence counts

Tumour Sample	RefSeq Gene	RefSeq ID	Sequence Count	Genomic Length	Chromosome	Sequence Orientation	Gene Orientation	Integration Locus	Intron Exon	% in Gene	Downstream of Gene
Adame T1	*Pah*	NM_008777	169	198	10	+	+	86961567	In2	16.00	0
Adame T1	*LOC382044*	NM_001081372	228	159	8	-	-	95966933	In2	9.50	0
Adame T1	*Acvr2a*	NM_007396	257	187	2	+	+	48100522			0
Gerry T1	*Setd5*	NM_028385	267	32	6	+	+	113057795	In1	18.97	0
Gerry T1	*Mrpl1*	NM_001039084	743	192	5	-	-	96434622			0
Jonas LT	*Rnf13*	NM_011883	10	131	3	+	+	57900377	In7	78.60	0
Jonas LT	*Uvrag*	NM_178635	29	1-114(115)	7	-	-	98830218	In12	73.16	0
Jonas LT	*Park7*	NM_020569	33	1-100(101)	4	-	-	149728872			12070

Integration sites in tumour derived EIAV infected mice. 454 sequencing was conducted by Manfred Schmidt's research group (Mannheim, Germany) on LAM PCR & non-restrictive PCR amplicons using the GS FLX system according to the manufactures instructions (Roche, Applied Science, USA). All sequences were conducted using Blas2seq & the Smith-Waterman algorithm. The raw sequences reads were aligned to the mouse genome using USL BLAT genome browser. Direct high sensitivity 454 pyrosequencing revealed several RefSeq genes with high sequence counts.

The genomic positions and orientation of provirus integration sites were retrieved using alignment tools such as BLAST and BLAT.

APPENDIX 11

The DNA sequence provided here is an '*in silico*' construction of the EIAV integration site flanking the *Pah* gene

MPAH Ex2-EIAV-Ex3

Ex2:25-132(108); In2 (1):133-3494; LAM PCR id: 3297-3494(198); 3'LTR: 3495-3575(g3539A); 5'LTR: 9113-9193; In2 (2):9194-15715 Ex3:15716-15899(184)

taaccaagtgatctaattatgtagGAAACAAGTTACATCGAAGACAACTCCAATCAAAATGGTG
CTGTATCTCTGATATT
CTCACTCAAAGAGGAAGTTGGTGCCCTGGCCAAGGTCCTGCGCTTATTTGAGgtaa
gtgtgtaatatatttcgatcacat
tcagccttcctctgatttctgcttgatctatttcttcttctctacccttaacatcaaactttgttgacttgccaggacca
cagcctcaaacaaacaaacaaacaaacaaacaaacaaaaacagaacaaaacaaaacaaaagaaccaacaacaacaacaac
aaaaaccaaactcccctcctctaccagcaattatatgacaatagcgcctcagctaagagtggcacttctggtctacctcc
actttctatagtgtagttttgcctggcttgagctccgcacaggtctcctgcctgctaccacaactgttttgagttcatat
gcataattgcactgctaggtctggaaagtgctgttttgctttagtcatcgttcctttggctcttacagttgtcccggttc
ttctgcaatgactctggcaatttccaaggaggggtgtgatacatacgactcattgaaggttgagcatgttagtccccctt
tctcgaaaacttgaccagctgtgggcttctattaaccaccatctaccacaaacagaagtttatctgataatagttgagag
atacactgatctgtaggcacgatagtgatacgttattaagtattggtttaatattatgtcaatttagctgaataaaagta
aggaggttctcccctcaaagcctatgacttgcagacagaccttcacaaacacatccagatgtttgttttctaggccattg
caaatccaatcaaattgataatgaagatcggccatcattatatgacattagatgcaattaacacaatatcaataacaatg
gcttggtactgtgtttaaaatccagagtgtttcacctgtcaggcaagtgctgtaatgctgagagacaccttcagccctat
ggctccatttcagacatctaaatctgctctagtaaagggagtgtatttagaacttaaaacccagttttctaaagtggtgg
tgatagttgtgggtaaatgaatgcctaatagcgagtagaaggcagtggggagaaaaaaaaggtgttttctgagggacaaa
ctgctcctgaaaatgatgtttttgtaaaatactccagtgcacttgcttagcctttgtatcagggaatcacatattcttga
gcagttgccccttcggactgagttagttaatggattaaataataataaatgaagaaactgttgacaatataataaacaaa
ccacaagcagaactggatgaagtgcacctaagaggtgtgtgtgtgcatgtttgtgcatgtttgtgtgtgtatttgtgtgt
gtgtgtacattcatgcatgcatgtatgaactctccatgaattctttgatacattatccagatctcaattccttgaggatc
tcactgtacccatcctgcagaggaagatatggtactcagcatgctgattggttggtggaaggctgttattagtgtatatt
gattttttctccaaaacaccctcagcagctatcatggaaactccccatcctggtacagcttttatacccttattaacag
cctgataaaattcatagtactataactccaagtgagcaagttaatcaagaaagcaaattcttctacacatggcaccacac
atatatgccaggcataaatcaggctcttgatttgatatccatgaaactcagaggtggaagccattcttttactgtatgga
ggagggagctgaggctccaagaaatgttgaacaattcactgaagataagtcagggaatgtgcaggcagaagtgggtttaa
aaccccgaaccccaggcttccaagtccatgatctgtctgtaacactcacccttctgccaaacaaagggagcttggcagcg
ggaggaagaggggaagctgctgaggtcactgtgcatctgtgtgtttctatcttttattacaccatcatgttctgacaaag
tttctggacataaaaagtgaaacttaaaagagtaaataacagagcaaatgaaaaaaaaagactcaactgtacaataaggc
aatatgatctttaaaaaacattctagcaaattctatattgttaaaaaagcattctgacaatccattctacatctggattt
tggagagccatagtgcacacacagacacttgcatgaccaggaagcctgcagtggaaagacccactcaatgtggtcacact
ttatatttcttaaacacatttgaccatggtaccttcttttttttttaattttttttattaggtattttcctcgtttacatt
ttcaatgctatcccaaaggtcccccatacccacccccccaatcccctacccacccactgtccctttttgaccctggtgtt
cccctgtactggggcatataaagtttgcaagtccaatgggcctctctttgcagtgatgaccgactaggccatcttttgat
acatatgcagctaaagacaagagctcccgggtactggttagttcatattgttgttccacctatagggttgcagttccctt
tagctccttgggtaatttctctagctcctccataaggggccgtgtgacccatccaatagctgactgtgatcatccacttc
tgtgtttgctaggccccggcatagtctcacaagagtgagctatatctggtcctttcagcaaccatggtaccttcttaaca
tctaggtttcagtattgtggtggcttaagttttgagctctaaactatcagcaacagcatgtgtggtgtgaaagacttttc
ctggggagacataactcacatagacaatgcagacctcttagcacatgaaagaccaaagtaaagactgcaacaaaggaatt
tttgttggggtcactatatcaagtacgggggaggtgttagttacaggagcaggaaagcaggaatgatttaaaaaaaaaaa

agcagtcaagccattgcataaccaagcaaaaaacatcctgggcaaatcatatgagtgacagctctcaaaacttacaatcc
tggagttcactgcaaaacttcctggtgatcactgttccaggataccaggctctttgcaacttagtaggtcagagcactct
cctagcacagtggatgctataaacctgaggggggaccttttagaatcttctgagttgtcactttccaggtcatgtgactct
ttgtttccctctagagTTTCCAGAGCCTCCTGCTCCCTCCTAGATAGAATCTTTCAGTTTGG
AAAAACTGCCATTCAGCA
GCCTGAATTTGTGGCACTGAGGAATAGAACCAAAGTGAAGAGTCTATTGTTTAG
AACCAAACTCAGTTGATCCACTTGTT
ATCGAGATGGGCAAAAGGAAGCTTTTAACCTCTTGGCTTAGTTTCCTTAGTTTcT
GTGGGGTTTTTATGAGGGGTTTTAT
AAATGATTATAAGAGTAAaAAGAAAGTTGCTGATGCTCTCATAACCTTGTATAA
CCCAAAGGACTAGCTCATGTTGCTAG
GCAACAATTGGGCACTCAGATTCTGCGGTCTGAGTCCCTTCTCTGCTGGGCTGAA
AAGGCCTTTGTAATAAATATAATTC
TCTACTCAGTCCCTGTCTCTAGTTTGTCTGTTCGAGATCCTACAgttggcgcccgaacagg
gacctgagaggggcgcaga
ccctacctgttgaacctggctgatcgtaggatccccgggacagcagaggagaacttacagaagtcttctggaggtgttcc
tggccagaacacaggaggacaggtaagattgggagaccctttgacattggagcaaggcgctcaagaagttagagaaggtg
acggtacaagggtctcagaaattaactactggtaactgtaattgggcgctaagtctagtagacttatttcatgataccaa
ctttgtaaaagaaaaggactggcagctgagggatgtcattccattgctggaagatgtaactcagacgctgtcaggacaag
aaagagaggcctttgaaagaacatggtgggcaatttctgctgtaaagatgggcctccagattaataatgtagtagatgga
aaggcatcattccagctcctaagagcgaaatatgaaaagaagactgctaataaaaagcagtctgagccctctgaagaata
tctctagagtcgacgctctcattacttgtaacaaagggagggaaagtatgggaggacagacaccatgggaagtatttatc
actaatcaagcacaagtaatacatgagaaacttttactacagcaagcacaatcctccaaaaaattttgtttttacaaaat
ccctggtgaacatggtcgactctagaactagtggatcccccgggctgcaggagtggggaggcacgatggccgctttggtc
gaggcggatccggccattagccatattattcattggttatatagcataaatcaatattggctattggccattgcatacgt
tgtatccatatcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattgattattgactagt
tattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatgg
cccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatggga
ctttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagt
acgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctac
ttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagc
ggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggac
tttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcag
agctcgtttagtgaaccgtcagatcgcctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat
ccagcctccgcggccccactagaactagtggatccgtcgaccatgcagcgcgtgaacatgatcatggcagaatcaccagg
cctcatcaccatctgccttttaggatatctactcagtgctgaatgtacagtttttcttgatcatgaaaacgccaacaaaa
ttctgaatcggccaaagaggtataattcaggtaaattggaagagtttgttcaagggaaccttgagagagaatgtatggaa
gaaaagtgtagttttgaagaagcacgagaagtttttgaaaacactgaaagaacaactgaattttggaagcagtatgttga
tggagatcagtgtgagtccaatccatgtttaaatggcggcagttgcaaggatgacattaattcctatgaatgttggtgtc
cctttggatttgaaggaaagaactgtgaattagatgtaacatgtaacattaagaatggcagatgcgagcagttttgtaaa
aatagtgctgataacaaggtggtttgctcctgtactgagggatatcgacttgcagaaaaccagaagtcctgtgaaccagc
agtgccatttccatgtggaagagtttctgtttcacaaacttctaagctcacccgtgctgagactgtttttcctgatgtgg
actatgtaaattctactgaagctgaaaccattttggataacatcactcaaagcacccaatcatttaatgacttcactcgg
gttgttggtggagaagatgccaaaccaggtcaattcccttggcaggttgttttgaatggtaaagttgatgcattctgtgg
aggctctatcgttaatgaaaaatggattgtaactgctgcccactgtgttgaaactggtgttaaaattacagttgtcgcag
gtgaacataatattgaggagacagaacatacagagcaaaagcgaaatgtgattcgaattattcctcaccacaactacaat
gcagctattaataagtacaaccatgacattgcccttctggaactggacgaacccttagtgctaaacagctacgttacacc
tatttgcattgctgacaaggaatacacgaacatcttcctcaaatttggatctggctatgtaagtggctggggaagagtct
tccacaaagggagatcagctttagttcttcagtaccttagagttccacttgttgaccgagccacatgtcttcgatctaca
aagttcaccatctataacaacatgttctgtgctggcttccatgaaggaggtagagattcatgtcaaggagatagtggggg
accccatgttactgaagtggaagggaccagtttcttaactggaattattagctggggtgaagagtgtgcaatgaaaggca
aatatggaatatataccaaggtatcccggtatgtcaactggattaaggaaaaaacaaagctcacttaatgaaagatggat
ttccaaggttaattcattggaattgaaaattaacagggcctctcactaactaatcactttcccatcttttgttagatttg
aatatatacattctatgatcattgctttttctctttacaggggagaatttcatattttacctgagcaaattgattagaaa

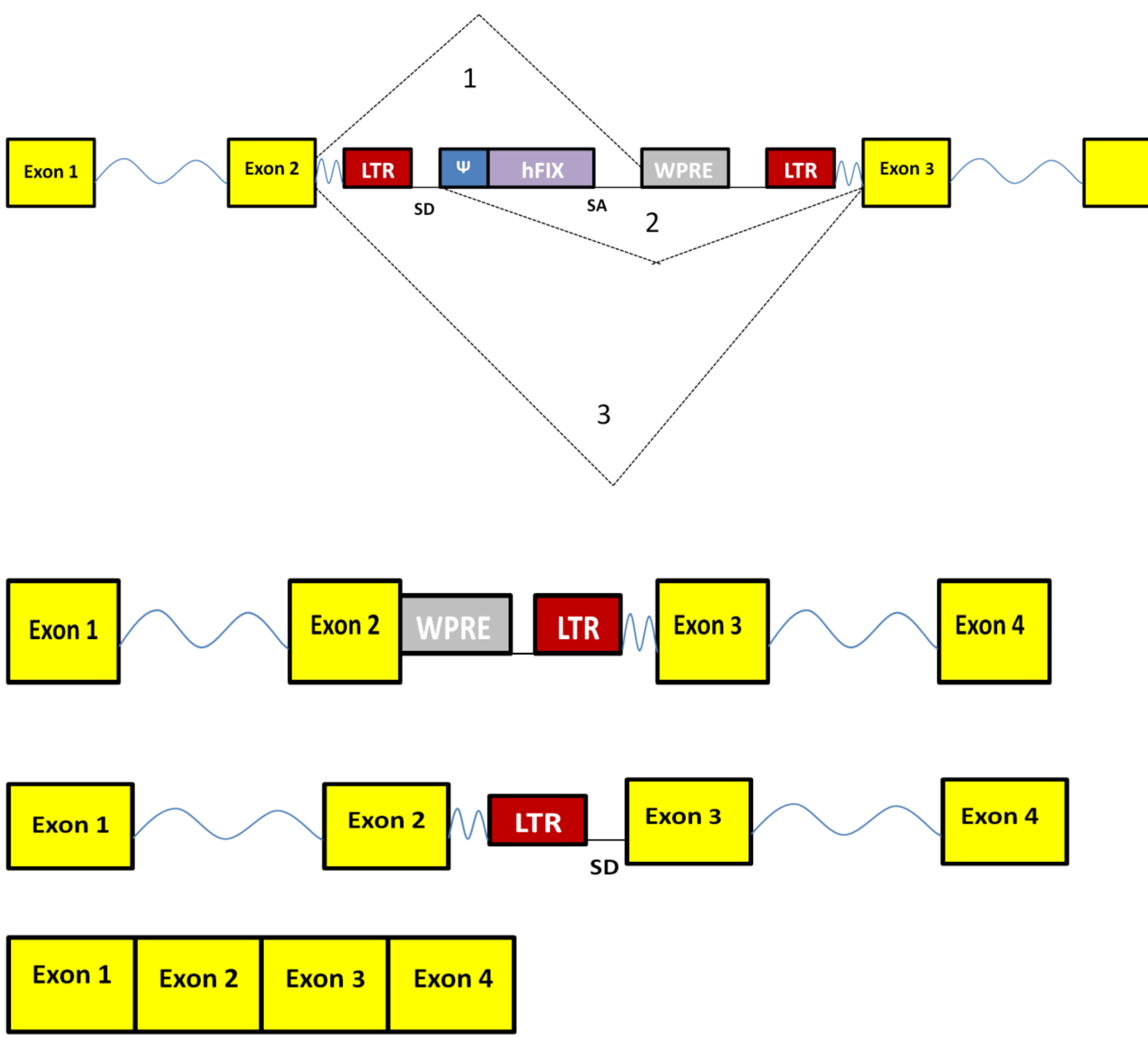
atggaaccactagaggaatataatgtgttaggaaattacagtcatttctaagggcccagcccttgacaaaattgtgaagt
taaattctccactctgtccatcagatactatggttctccactatggcaactaactcactcaattttccctccttagcagc
attccatcttcccgatcttctttgcttctccaaccaaaacatcaatgtttattagttctgtatacagtacaggatccccc
gggctgcaggaattcgatatcaagcttatcgataccgtcgacctcgagggggggcccgacctactagggtgctgtggaag
ggtgatggtgcagtagtagttaatgatgaaggaaagggaataattgctgtaccattaaccaggactaagttactaataaa
accaaattgagtattgttgcaggaagcaagacccaactaccattgtcagctgtgtttcctgacctcaatatttgttataa
ggtttgatatgaatcccaggggaatctgaacccctattacccaacagtcagaaaaatctaagtgtgaggagaacacaat
gtttcaaccttattgttataataatgacagtaagaacagcatggcagaatcgaaggaagcaagagaccaagaatgaacct
gaaagaagaatctaaagaagaaaaagaagaaatgactggtggaaaataggtatgtttctgttatgcttagcaggaacta
ctggaggaatactttggtggtatgaaggactcccacagcaacattatatagggttggtggcgataggggaagattaaac
ggatctggccaatcaaatgctatagaatgctggggttccttcccggggtgtagaccatttcaaaattacttcagttatga
gaccaatagaaggtgaccagtggtgcagggtcctccggcagtcgttacctgaagaaaaaattccatcacaaacatgcatc
gcgagaagacacctgggaccaggcccaacacaacatacacctagcaggcgtgaccggtggatcaggggacaaatactaca
agcagaagtactccaggaacgactggaatggagaatcagaggagtacaacaggcggccaaagagctgggtgaagtcaatc
gaggcatttggagagagctatatttccgagaagaccaaaggggagatttctcagcctggggcggctatcaacgagcacaa
gaacggctctgggggaacaatcctcaccaagggtccttagacctggagattcgaagcgaaggaggaaacatttatgact
gttgcattaaagcccaagaaggaactctcgctatcccttgctgtggatttcccttatggctattttgggggctagtaatt
atagtaggacgcatagcaggctatggattacgtggactcgctgttataataaggatttgtattagaggcttaaatttgat
atttgaaataatcagaaaaatgcttgattatattagcttcgaatcactagtgaattcgcggccgcaatcaacctctggat
tacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgc
ctttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgag
gagttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgc
caccacctgtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttg
cccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcggggaagctgacgtcctttccatgg
ctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggacct
tccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatctccctttt
gggccgcctccccgcctgatcgatgtcgagaatcactagtgaattcgcggccgctcgacagcttatcgataccgtcgaat
tggaagagctttaaatcctggcacatctcatgtatcaatgcctcagtatgtttagaaaaacaagggggggaacTGTGGGGT
TTTTATGAGGGGTTTTATAAATGATTATAAGAGTAAaAAGAAAGTTGCTGATGCT
CTCATAACCTTGTATAACccaaagg
actagctcatgttgctaggcaacaattgggcactcagattctgcggtctgagtcccttctctgctgggctgaaaaggcct
ttgtaataaatataattctctactcagtccctgtctctagtttgtctgttcgagatcctacAaaagtgaaggtagtggta
aaccttcaccaatagggtttccacaaaaatcaaaccaaatcatttgtgaaaactaattataacagcaactgtcatagagt
aagttctcaattagcattttttttcaaaagttgttgctgtgtactgtagcctaaagaactcttaaattcttggatatagg
aaaacctgttgaagactgaggttcgtttggttggaaaggaaaagactctgtatgggaatgctcacctgtttgaggttgtc
catggaaggaaacttctagatatttgttagtgtcaagacctcaatgagtagaaataccgggaatatacttcacatcattt
acaagtgattctttaatatagaggacaatgggaggcgctcatttcaagtatcccttctgttaagttaaacccccaaagaa
caacatcagagtaaaaactgtgatctaaatacacttttattttttttttcttttacagaaggtaaaaatctatattctca
atattggtctgagagggaaaagtatttcatgtggggataatcacctatctactgatacttttcttattaaatagccttat
aatgactctaggcttaatgagcagccccaacactcaatacttatagcatggcagtcatacaactggataattagccaact
cttactaaatactcagaggactaaggagagactcacagtgatgtacagtctctagactagaatgcgaagggtcagactag
atcaccttgtcctctggcctgctgaaagtggacactttcttgacttggctgctagtagaacatactaaagcatgcaggca
tctcagagaaagaaaggcctgtagtatgtgtgtaggcccctcttggattagtcaggattacgtaccttccattcaacttc
aaagccaccacctcataaaggaaccaatttttttctctttgggcttgtagacttcagctggtaactacatcagaatgca
taccaatgtagcccatatttatttgtttttgtttcattttccagtaatttatttttattagctttgcatcctgatcatag
cctcctctctcctcagtcccacttttacaaatcccttccccccaataccccctgcccttctcctcagagatagggagcccc
cccttgggtaccatcctatccgggacataaagtctcagcgggactctgcacatcctctctcactgaggcccaaccaggaa
gtccaggtcgggagaagggaatccaatggctgggaataagcaaagaaacagatccacaaaaacccgagttgtgcgaggtc
atctcagtcattctccagcagcctcatctagactgtttaatcggtggcatcgaagagcttcaggacaagcaagttgccaa
catctggagaggctcagctggatattgcttctgaatttatactccactttcgtggatatgattcatcttttagaaacaat
tgaaaggaaactgactacatttattacgcagatttcctgtgccctactctcgtccagggaacacactcctgccacccctc
cactgttggatgtagattctgctgttaggttctatagagacgatggagtcactgaagtgagctagtcatccctccttggt
ggctatagcacttagactccttgctaccaggcaatacaaatagagttatcattgttaacttgtgttacgaaagctgttgt
tacttgctttataggattttattggatgattgtgctataatccatttaaccattccccagctgagagtgtttaattagat

gacaaaatttggttgaaagactttgtgaattttttccctgagttagatttccagtattataacaaatgaagtcattctg
tataaactgttgttttgtggttattcccaaagcaaggagctggtactgcttccgcttcttctcttctccccctcttcctc
ctcttctttttcctcctcctcttcctcctctttctcccccatctcttcctcttcctcctccttttcctcttattctcccc
tcctccttccccatcttctcttcctttctcttctccttttgttttttattgggatggtgtctcattagatagctttagct
gtcctggggctcatgtagagcaggttggcttcatacagaggcctcctgcctctacctccctagtatgctaccacttctgg
agtacaaagagcttcttggagaaagtggctacttccagggatgtcttatgggtgggtggtcaagttgtcctaaagtgtgg
tttgccttctggcatgtcttctaccatgtgcacacacatcttctcttctgtccttgtaccttctcaaacggttagaca
aatacatgagctcaccttgcttcttggctttactgtatccctttctctgatttttgccacatacataaactttcttgcca
ttatttattaatttttatcatgcagtatattttgatcacatttccccatcccctcatgcctcgcagaccctccccacttt
gctctatacagcatagaccttggttgcctcttaaattttaacagaagtcaataggggaggggatgtgaccccaggaaggg
ttcacaaattctctttcttcctaagccaagaagcatgaccttgctttttgctttgtttaggaaatgacctctgtttgctc
atagtagcttttgagttccagctgtctctctggtcctccctctgcagtatttctctgtcctgcttggggttttgatgttt
cccacaccagatcagggttctctttgtctgcagatcaggatatagctctcagctgctgttccagcatatgctgccaaacc
tctgaaactataagcaagcttctaatttaaatgctttcttttataagagttgcattggtcatgatgtctgtcttagtcag
ggtttctattcctgcacaaacatcatgaccaagaagcaagttagggaggaaagggtttattcagcttacatttccatatt
gctgttcatcaccaaaggatgcaggactggaactcaagcaggtcaggaagcaggagctgatgcagaggccatggagggat
gttccttactagcttgcctctcctggcctcctcagcctgctctcttatcgaaccaagactaccagccccgagatcgtccc
acccacaaggggcctttccccccttgatcactaattgagaaaatgccttacagctggatctcatggaggcatttcctcaac
tgaagctcctttctctgtgataactccagctgtgtcaagttgacacaaaactagccagtacaatgtctcttcatagtaat
agaacagtgactaagacccacaatcatacacatatgcacagaatcatttaagggaagctgtaaccagtggtaaaatcctg
gggcctcaggggacacaaagaagtctcggcagaaagagagcatgtgtcaggttcattactctgttccttgtacccagaga
ttctgatgcttttgtgtttgatgtcatattaatgataatgaaaaaaaatcaacattttaagagagcttactgtatccctg
gaagagaaggcagtgtttacaaaagaaaaatgactttgctgtgaatcatgtgctttcttcctctgttcttcacaacagag
ctttttcatctatgattttattgtttctcttcctcaatttcctcaaaggccaaaaagaagggcccataaatgaccttcat
aggtggtatttgagaggaggagctatcaccattgtagtgacatatgttctcctcagcattgcattgtttataactggaca
tacactgcttcaaccacaaaaagggtgaattttattttaagttttggatctattcaagttcctgcacctgcagtacaagc
tttgtccatttgaatattggtatttattttttaagttatgaattcatgatgtgtcctggttttctacttagtgaacact
aacatagctgtactggtggaatcagtcatagtgtaaagcatatgttttttaaagtatttttaaatttgggagcccatgta
agtgattctatatgtaggtgtttgtgtgtgtgtcttagtcattgttatttggcatgaagagacaccatgacaaatgcagc
tcttatgaaagaaagcatttagttagaagcttgcataaagtttcagaggtttggtctaatatcatcatggcagggaacat
gacagcatatgctggatcagtagctgacagctacatcctgaaatgaaggcccagagggagaacctgggcctggtgtcggc
ttttgaaacatcaaagcttaaaaattgtttgtgtgaagatatttaaaaatagacaaggatagtattatcataaaaaataa
tacacagtagaattaatatcaatctatacacaagtattactgatcagaaaatagaatataggttctctcatggagggtag
aacagatgatgttgccccttttgagagcagcggggtaagtaagatacagaaactttaaggggtgaatactttctggttcac
catggaggccccaaggcatttaatttgggaaatcatctaggtaatcagggacaagcatgcagacaatgtcagtcattgct
tctagtggcaaagcatatgaggaaagtcacttgccagtcactaaaacactggttgtagaaatcaaggcacgtctagatag
tgtactctagcaagtgctgaaaaattgtataatgggtccgctggcagattgaatctgcaggtaaaggcattttctgccaa
gtctgatgacttgtttttatctctgagatccacagctatggtagatgtggctaatgacaaataagaagtatttgttcatc
tctctagctttgctatagatttttttctttttttaattcaaagaagaaactctagggagatttttgtgtatgtgctggcc
ccttaggaattctgaattctcactcagctgcacagtgtttctaaacagacgatgaaatgggacctactgatccctttccc
attttgttatctacaaattcctctcatagtccttatcattttactctttgaggtcattacatgaaatatcagcctgtgac
tcctgacgctgtcacacttgccaccatagttacacccactatcatatgggatttcagaagcaactttgcggcagaactta
gtggttctgccaggtctacaagggttattctcaacatggatgacagattgatcattctaataccatgagatccggttttc
tgtctgctcaaaacttcataaggaccccctaaaagtttcagaatcaaggccagatccattcaaataatttgtctagctgtc
cttcatcttttgtggctattccaacttatcagacttcttttattacttcttctaggacaatcccatccaccattcccatg
gggtcatcctgtctgttaactcctttagaatgtgcatatggtttgctttttcaagaagacctcatgtgcagatcctctta
aatactagactgtatcctaatgctaccccagggcagctcatccattttgtctgtcctgccttgtatttatttttatcccc
tgacattataaaataatacttaagttaattgctttttatggctgctgtctttattacctcaacttgaatgtaaactttag
aaagaaaaaaacatctaggtctgagttgtgattggcatttagtgaatactgcatatgtatttatataggttatatatgta
tatgtacatgcattttcaaatataatacaaataatataccttacatagttatataacatatgatgatattaatagtagtt
ctacaaattatgttgcagagtgtatgtatgtattaatttgtccaagacccccaaagagctcatggtgcagttattaatga
gaccacccatgtgtagatggggaaactgaggctatgcagaagcatgaaaataaaatagctagcatttcatagatgataca
atttagagattatgaatgcagagcctgtgccctggggattatattcaagcaaaacaatgttgtgtctatatgatttattg
atttctgtggcatttcattgggtggaagggagactttcccatgccatacctttacaaccattgtcttagacatttgatcc

aaacccaagccttacacaggcaactgtaggctggtattagatttctgtgtgattaagtttggatgcacttgactcattct
tcctatctcttgtcaattggcgcttggttttgcagaaggtaaaacagagtctaaatatttaagggtttccttattctcat
gtaatacaaaatgtttttaaaccctgctcttcaactctagttgttctggctctgtatatttggactgtggtaataatgaa
caccagactcccacacctataattttctctttcagGAGAATGAGATCAACCTGACACACATTGAATCC
AGACCTTCCCGT
TTAAACAAAGATGAGTATGAGTTTTTCACCTATCTGGATAAGCGTAGCAAGCCC
GTCCTGGGCAGCATCATCAAGAGCCT
GAGGAACGACATTGGTGCCACTGTCCATGAGCTTTCCCGAGACAAGGAAAAGA
ACACAGgtgagagtctgaagaatctcg
tgacctaa

APPENDIX 11

DNA sequences representing the potential splice combinations of the *Pah* gene



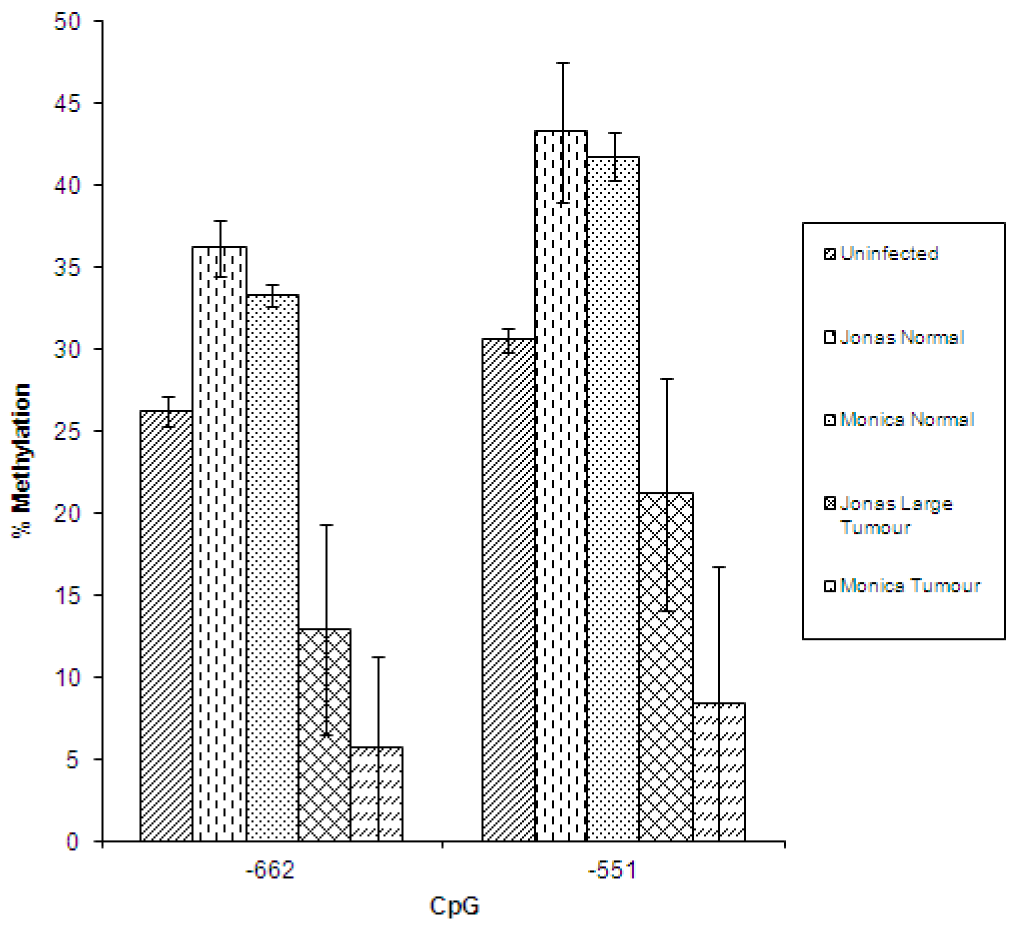
APPENDIX 12

Bisulphite sequences used for methyl specific PCR and DNA sequencing

Gene	Fragment	Forward Primer	Reverse Primer (5' Biotin)	Annealing Temperature (°C)
Hmgcs2	1	TGGTTTAGTTTTTTTTTAAGTAGG	TCACCTCTACTAAACTACTAAAAACAA	56
	2	TGAGGGATTTTGAATTTTTTTAGT	TTTTAATACCAACTCTTCAAACAATT	48.1
Park7	1	GTGGGGTTTAGGTTGTAGTTTAGTTA	TCATATCCCACTAAACCCAAAC	48.4
	2	TTATTTTTATTTAGGGGTTTGGGTTTA	AATCACAACCCCACCTCTAACC	48.1
	3	GGTTAGAGTTTTGGGGTTAGAGG	TCTCTTTAAAAAATCAACCCTACTAC	61

Gene	Fragment	Sequencing Primer	Dispensation Order
Hmgcs2	1	GTTTTTTTTTAAGTAGGTA TTATAAAGTTTAATTTAGAA GTTGATATTTTAGAGATGTA	CGGTCTGATAGTCTGAGTCTG ATCAGTGTCGATT AGTCAGTGTTAGTCTGAGTCAGTTGTAGTGAGAGAGTCTGATGTAGATCTG
	2	TTTTTTAGTAGTATTTGTTG AGTTTTGTGTAAGTATAATG TTTTATAGATTAAAGTTAGTTA	ATCAGTTGAGTTGTGAGTTGTGTAGTCGATATGATCTG GATCTGATCGAT CGTCAGTAGATGTGTCAGTATCGA
Park7	1	TTTAGTTAGGTTTTAAATT AGAGGTAGAGAGGGGTAAT GGGGGAGAGATGGTATAAG	ATCTGATATATCGATATCAGTCTG CTGTCAGTAGTCAGTCAGTAGTGTCAGTATCAGTGTGGTCTG ATCAGTGTCAGTAGTCAGTAGTAGTTATCAGTTATTATCTG
	2	ATTTAGGGGTTTGGGTTTAGT AGGGGTTGTGAGTAGTTTTAG GTATGAGTGTTGGGTATA GGTGAGAGGTTTTTGAGGATTAG	ATCTGATATGATCAGTAGTCTGAGTCTGAGTCTGAGTGTGTCAG CTGTCAGTGTCAGTCAGTATGTCTG ATCAGTCAGTGTCTGAGTTGTCTG ATCAGTGTCAGTCAGTGTGAGAGTCAGTCTGTGTAGTAGTCAG
	3	GTTTTGGGGTTAGAGGT TTGGGAGGGTTGGGGGTA GTTTGATTAAAGTTGTGATG	ATCAGTGTGATCAGTCAGTGTCTGAGTCAGTGTCTG AGTCAGTCAGTAGTGTATGTCAGTAGTTGTATAGTATCTGTGAGTCTGAGTCAG ATCAGTGGAGTCAGTGTAGAGTATCAGTCTGAGTGTCTG
		GGAAAGTGTTTTAGAGTGG	ATCAGTAGTGTCAGTC

APPENDIX 13



Adams, R.L.P. (1990) "DNA methylation. The effect of minor bases on DNA-protein interactions", *Biochemical Journal,* vol. 265, no. 2, pp. 309-320.

Aiuti, A., Cattaneo, F., Galimberti, S., Benninghoff, U., Cassani, B., Callegaro, L., Scaramuzza, S., Andolfi, G., Mirolo, M., Brigida, I., Tabucchi, A., Carlucci, F., Eibl, M., Aker, M., Slavin, S., Al-Mousa, H., Ghonaium, A.A., Ferster, A., Duppenthaler, A., Notarangelo, L., Wintergerst, U., Buckley, R.H., Bregni, M., Marktel, S., Valsecchi, M.G., Rossi, P., Ciceri, F., Miniero, R., Bordignon, C. and Roncarolo, M.-. (2009) "Gene therapy for immunodeficiency due to adenosine deaminase deficiency", *New England Journal of Medicine,* vol. 360, no. 5, pp. 447-458.

Aiuti, A., Slavin, S., Aker, M., Ficara, F., Deola, S., Mortellaro, A., Morecki, S., Andolfi, G., Tabucchi, A., Carlucci, F., Marinello, E., Cattaneo, F., Vai, S., Servida, P., Miniero, R., Roncarolo, M.G. and Bordignon, C. (2002a) "Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning", *Science,* vol. 296, no. 5577, pp. 2410-2413.

Aiuti, A., Vai, S., Mortellaro, A., Casorati, G., Ficara, F., Andolfi, G., Ferrari, G., Tabucchi, A., Carlucci, F., Ochs, H.D., Notarangelo, L.D., Roncarolo, M.G. and Bordignon, C. (2002b) "Immune reconstitution in ADA-SCID after PBL gene therapy and discontinuation of enzyme replacement [1]", *Nature medicine,* vol. 8, no. 5, pp. 423-425.

Akagi, K., Suzuki, T., Stephens, R.M., Jenkins, N.A. and Copeland, N.G. (2004) "RTCGD: Retroviral tagged cancer gene database", *Nucleic acids research,* vol. 32, no. DATABASE ISS., pp. D523-D527.

Albanese, A., Arosio, D., Terreni, M. and Cereseto, A. (2008) "HIV-1 pre-integration complexes selectively target decondensed chromatin in the nuclear periphery", *PLoS ONE,* vol. 3, no. 6, art no. e2413.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: A new generation of protein database search programs", *Nucleic acids research,* vol. 25, no. 17, pp. 3389-3402.

Ambrosi, A., Cattoglio, C. and Di Serio, C. (2008) "Retroviral integration process in the human genome: Is it really non-random? A new statistical approach", *PLoS Computational Biology,* vol. 4, no. 8, art no. e1000144.

Amsterdam, A., Burgess, S., Golling, G., Chen, W., Sun, Z., Townsend, K., Farrington, S., Haldi, M. and Hopkins, N. (1999) "A large-scale insertional mutagenesis screen in zebrafish", *Genes and Development,* vol. 13, no. 20, pp. 2713-2724.

Anderson, J.L. and Hope, T.J. (2005) "Intracellular trafficking of retroviral vectors: Obstacles and advances", *Gene therapy,* vol. 12, no. 23, pp. 1667-1678.

Avila, M.A., Berasain, C., Sangro, B. and Prieto, J. (2006) "New therapies for hepatocellular carcinoma", *Oncogene,* vol. 25, no. 27, pp. 3866-3884.

Avilés Mendoza, G.J., Seidel, N.E., Otsu, M., Anderson, S.M., Simon-Stoos, K., Herrera, A., Hoogstraten-Miller, S., Malech, H.L., Candotti, F., Puck, J.M. and Bodine, D.M. (2001) "Comparison of five retrovirus vectors containing the human IL-2 receptor γ chain gene for their ability to restore T and B lymphocytes in the X-linked severe combined immunodeficiency mouse model", *Molecular Therapy,* vol. 3, no. 4, pp. 565-573.

Azzouz, M., Kingsman, S.M. and Mazarakis, N.D. (2004) "Lentiviral vectors for treating and modeling human CNS disorders", *Journal of Gene Medicine,* vol. 6, no. 9, pp. 951-962.

Azzouz, M. and Mazarakis, N. (2004) "Non-primate EIAV-based lentiviral vectors as gene delivery system for motor neuron diseases", *Current Gene Therapy,* vol. 4, no. 3, pp. 277-286.

Babior, B.M. (2004) "NADPH oxidase", *Current opinion in immunology,* vol. 16, no. 1, pp. 42-47.

Bainbridge, J.W.B. (2009) "Prospects for gene therapy of inherited retinal disease", *Eye,* vol. 23, no. 10, pp. 1898-1903.

Balaggan, K.S., Binley, K., Esapa, M., Iqball, S., Askham, Z., Kan, O., Tschernutter, M., Bainbridge, J.W.B., Naylor, S. and Ali, R.R. (2006) "Stable and efficient intraocular gene transfer using pseudotyped EIAV lentiviral vectors", *Journal of Gene Medicine,* vol. 8, no. 3, pp. 275-285.

Baltimore, D. (1970) "Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of RNA tumour viruses", *Nature,* vol. 226, no. 5252, pp. 1209-1211.

Bardwell, L. (1989) "The mutagenic and carcinogenic effects of gene transfer", *Mutagenesis,* vol. 4, no. 4, pp. 245-253.

Barraza, R.A. and Poeschla, E.M. (2008) "Human gene therapy vectors derived from feline lentiviruses", *Veterinary immunology and immunopathology,* vol. 123, no. 1-2, pp. 23-31.

Barry, A.E., Leliwa-Sytek, A., Tavul, L., Imrie, H., Migot-Nabias, F., Brown, S.M., McVean, G.A.V. and Day, K.P. (2007) "Population genomics of the immune evasion (var) genes of Plasmodium falciparum", *PLoS Pathogens,* vol. 3, no. 3, art no. 34.

Bartholomew, C. and Ihle, J.N. (1991) "Retroviral insertions 90 kilobases proximal to the Evi-1 myeloid transforming gene activate transcription from the normal promoter", *Molecular and cellular biology,* vol. 11, no. 4, pp. 1820-1828.

Bates, P., Young, J.A.T. and Varmus, H.E. (1993) "A receptor for subgroup A Rous sarcoma virus is related to the low density lipoprotein receptor", *Cell,* vol. 74, no. 6, pp. 1043-1051.

Baum, C. (2011) "Gene therapy for SCID-X1: Focus on clinical data", *Molecular Therapy,* vol. 19, no. 12, pp. 2103-2104.

Baum, C., Düllmann, J., Li, Z., Fehse, B., Meyer, J., Williams, D.A. and Von Kalle, C. (2003) "Side effects of retroviral gene transfer into hematopoietic stem cells", *Blood,* vol. 101, no. 6, pp. 2099-2114.

Baum, C., von Kalle, C., Staal, F.J.T., Li, Z., Fehse, B., Schmidt, M., Weerkamp, F., Karlsson, S., Wagemaker, G. and Williams, D.A. (2004) "Chance or necessity? Insertional mutagenesis in gene therapy and its consequences", *Molecular Therapy,* vol. 9, no. 1, pp. 5-13.

Bedford, M.T. and Van Helden, P.D. (1987) "Hypomethylation of DNA in pathological conditions of the human prostate", *Cancer research,* vol. 47, no. 20, pp. 5274-5276.

Bendich, A. and Rosenkranz, H.S. (1961) "A correlation between the sedimentation coefficient of deoxyribonucleic acid and its paper chromatographic migration", *Archives of Biochemistry and Biophysics,* vol. 94, no. 3, pp. 417-423.

Bendinelli, M., Pistello, M., Lombardi, S., Poli, A., Garzelli, C., Matteucci, D., Ceccherini-Nelli, L., Malvaldi, G. and Tozzini, F. (1995) "Feline immunodeficiency virus: An interesting model for AIDS studies and an important cat pathogen", *Clinical microbiology reviews,* vol. 8, no. 1, pp. 87-112.

Bennet, M. and Smyth, N.R. (1992) "Feline immunodeficiency virus. A brief review", *British Veterinary Journal,* vol. 148, no. 5, pp. 399-412.

Berger, S.L. (2007) "The complex language of chromatin regulation during transcription", *Nature,* vol. 447, no. 7143, pp. 407-412.

Bernstein, E. and Hake, S.B. (2006) "The nucleosome: A little variation goes a long way", *Biochemistry and Cell Biology,* vol. 84, no. 4, pp. 505-517.

Biasco, L., Ambrosi, A., Pellin, D., Bartholomae, C., Brigida, I., Roncarolo, M.G., Di Serio, C., Von Kalle, C., Schmidt, M. and Aiuti, A. (2011) "Integration profile of retroviral vector in gene therapy treated patients is cell-specific according to gene expression and chromatin conformation of target cell", *EMBO Molecular Medicine,* vol. 3, no. 2, pp. 89-101.

Bild, A.H., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.-., Harpole, D., Lancaster, J.M., Berchuck, A., Olson Jr., J.A., Marks, J.R., Dressman, H.K., West, M. and Nevins, J.R. (2006) "Oncogenic pathway signatures in human cancers as a guide to targeted therapies", *Nature,* vol. 439, no. 7074, pp. 353-357.

Billard, L.-., Magdinier, F., Lenoir, G.M., Frappart, L. and Dante, R. (2002) "MeCP2 and MBD2 expression during normal and pathological growth of the human mammary gland", *Oncogene,* vol. 21, no. 17, pp. 2704-2712.

Bird, A.P. (1986) "CpG-rich islands and the function of DNA methylation", *Nature,* vol. 321, no. 6067, pp. 209-213.

Bird, A.P. and Southern, E.M. (1978) "Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis", *Journal of Molecular Biology,* vol. 118, no. 1, pp. 27-47.

Bird, A.P. and Wolffe, A.P. (1999) "Methylation-induced repression-belts, braces, and chromatin", *Cell,* vol. 99, no. 5, pp. 451-454.

Bishop, J.M. (1983) "Cellular oncogenes and retroviruses", *Annual Review of Biochemistry,* vol. Vol. 52, pp. 301-354.

Björgvinsdóttir, H., Zhen, L. and Dinauer, M.C. (1996) "Cloning of murine gp91phox cDNA and functional expression in a human X-linked chronic granulomatous disease cell line", *Blood,* vol. 87, no. 5, pp. 2005-2010.

Black, D.L. (2003) "Mechanisms of alternative pre-messenger RNA splicing", *Annual review of biochemistry,* 2003, vol. 72, pp. 291-336.

Blackburn, M.R., Aldrich, M., Volmer, J.B., Chen, W., Zhong, H., Kelly, S., Hershfield, M.S., Datta, S.K. and Kellems, R.E. (2000) "The use of enzyme therapy to regulate the metabolic and phenotypic consequences of adenosine deaminase deficiency in mice. Differential impact on pulmonary and immunologic abnormalities", *Journal of Biological Chemistry,* vol. 275, no. 41, pp. 32114-32121.

Blackburn, M.R., Datta, S.K., Wakamiya, M., Vartabedian, B.S. and Kellems, R.E. (1996) "Metabolic and immunologie consequences of limited adenosine deaminase expression in mice", *Journal of Biological Chemistry,* vol. 271, no. 25, pp. 15203-15210.

Blaese, R.M., Culver, K.W., Miller, A.D., Carter, C.S., Fleisher, T., Clerici, M., Shearer, G., Chang, L., Chiang, Y., Tolstoshev, P., Greenblatt, J.J., Rosenberg, S.A., Klein, H., Berger, M., Mullen, C.A., Ramsey, W.J., Muul, L., Morgan, R.A. and Anderson, W.F. (1995) "T lymphocyte-directed gene therapy for ADA- SCID: Initial trial results after 4 years", *Science,* vol. 270, no. 5235, pp. 475-480.

Blais, A. and Dynlacht, B.D. (2004) "Hitting their targets: An emerging picture of E2F and cell cycle control", *Current Opinion in Genetics and Development,* vol. 14, no. 5, pp. 527-532.

Boerkoel, C.F. and Kung, H.-. (1992) "Transcriptional interaction between retroviral long terminal repeats (LTRs): Mechanism of 5' LTR suppression and 3' LTR

promoter activation of c- myc in avian B-cell lymphomas", *Journal of virology,* vol. 66, no. 8, pp. 4814-4823.

Boonyaratanakornkit, J., Chew, A., Ryu, D.D.Y., Greenhalgh, D.G. and Cho, K. (2004) "Murine endogenous retroviruses and their transcriptional potentials", *Mammalian Genome,* vol. 15, no. 11, pp. 914-923.

Borenfreund, E., Fitt, E. and Bendich, A. (1961) "Isolation and properties of deoxyribonucleic acid from mammalian sperm", *Nature,* vol. 191, no. 4796, pp. 1375-1377.

Bosticardo, M., Ghosh, A., Du, Y., Jenkins, N.A., Copeland, N.G. and Candotti, F. (2009) "Self-inactivating retroviral vector-mediated gene transfer induces oncogene activation and immortalization of primary murine bone marrow cells", *Molecular Therapy,* vol. 17, no. 11, pp. 1910-1918.

Bracken, A.P., Ciro, M., Cocito, A. and Helin, K. (2004) "E2F target genes: Unraveling the biology", *Trends in biochemical sciences,* vol. 29, no. 8, pp. 409-417.

Brady, T. and Bushman, F.D. (2011) "Nondividing cells: A safer bet for integrating vectors?", *Molecular Therapy,* vol. 19, no. 4, pp. 640-641.

Brandt, J., Schrauth, S., Veith, A.-., Froschauer, A., Haneke, T., Schultheis, C., Gessler, M., Leimeister, C. and Volff, J.-. (2005) "Transposable elements as a source of genetic innovation: Expression and evolution of a family of retrotransposon-derived neogenes in mammals", *Gene,* vol. 345, no. 1 SPEC. ISS., pp. 101-111.

Bremner, R. and Zacksenhaus, E. (2010) "Cyclins, Cdks, E2f, Skp2, and more at the first international RB tumor suppressor meeting", *Cancer research,* vol. 70, no. 15, pp. 6114-6118.

Brenner, S. and Malech, H.L. (2003) "Current developments in the design of onco-retrovirus and lentivirus vector systems for hematopoietic cell gene therapy", *Biochimica et Biophysica Acta - Molecular Cell Research,* vol. 1640, no. 1, pp. 1-24.

Brenner, S., Ryser, M.F., Choi, U., Whiting-Theobald, N., Kuhlisch, E., Linton, G., Kang, E., Lehmann, R., Rosen-Wolff, A., Rudikoff, A.G., Farese, A.M., MacVittie, T.J., Roesler, J., Horwitz, M.E. and Malech, H.L. (2006) "Polyclonal Long-Term MFGS-gp91phox Marking in Rhesus Macaques after Nonmyeloablative Transplantation with Transduced Autologous Peripheral Blood Progenitor Cells", *Molecular Therapy,* vol. 14, no. 2, pp. 202-211.

Brenner, S., Whiting-Theobald, N.L., Linton, G.F., Holmes, K.L., Anderson-Cohen, M., Kelly, P.F., Vanin, E.F., Pilon, A.M., Bodine, D.M., Horwitz, M.E. and Malech, H.L. (2003) "Concentrated RD114-pseudotyped MFGS-gp91phox vector achieves high levels of functional correction of the chronic granulomatous disease oxidase defect in NOD/SCID/β2-microglobulin-/-

repopulating mobilized human peripheral blood CD34+ cells", *Blood,* vol. 102, no. 8, pp. 2789-2797.

Brielmeier, M., Béchet, J.-., Falk, M.H., Pawlita, M., Polack, A. and Bornkamm, G.W. (1998) "Improving stable transfection efficiency: Antioxidants dramatically improve the outgrowth of clones under dominant marker selection", *Nucleic acids research,* vol. 26, no. 9, pp. 2082-2085.

Brolin, C. and Shiraishi, T. (2011) "Antisense mediated exon skipping therapy for duchenne muscular dystrophy (DMD)", *Artificial DNA: PNA and XNA,* vol. 2, no. 1, pp. 6-15.

Brown, K.D. and Robertson, K.D. (2007) "DNMT1 knockout delivers a strong blow to genome stability and cell viability", *Nature genetics,* vol. 39, no. 3, pp. 289-290.

Buchschacher Jr., G.L. and Wong-Staal, F. (2000) "Development of lentiviral vectors for gene therapy for human diseases", *Blood,* vol. 95, no. 8, pp. 2499-2504.

Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S. and Hoffmann, C. (2005) "Genome-wide analysis of retroviral DNA integration", *Nature Reviews Microbiology,* vol. 3, no. 11, pp. 848-858.

Bushman, F.D. (2003) "Targeting survival: Integration site selection by retroviruses and LTR-Retrotransposons", *Cell,* vol. 115, no. 2, pp. 135-138.

Bushman, F.D., Hoffmann, C., Ronen, K., Malani, N., Minkah, N., Rose, H.M., Tebas, P. and Wang, G.P. (2008) "Massively parallel pyrosequencing in HIV research", *AIDS,* vol. 22, no. 12, pp. 1411-1415.

Callahan, R. and Smith, G.H. (2000) "MMTV-induced mammary tumorigenesis: Gene discovery, progression to malignancy and cellular pathways", *Oncogene,* vol. 19, no. 8, pp. 992-1001.

Calmels, B., Ferguson, C., Laukkanen, M.O., Adler, R., Faulhaber, M., Kim, H.-., Sellers, S., Hematti, P., Schmidt, M., Von Kalle, C., Akagi, K., Donahue, R.E. and Dunbar, C.E. (2005) "Recurrent retroviral vector integration at the Mds1/Evi1 locus in nonhuman primate hematopoietic cells", *Blood,* vol. 106, no. 7, pp. 2530-2533.

Calzone, L., Gelay, A., Zinovyev, A., Radvanyl, F. and Barillot, E. (2008) "A comprehensive modular map of molecular interactions in RB/E2F pathway.", *Molecular systems biology,* vol. 4, pp. 173.

Cam, H. and Dynlacht, B.D. (2003) "Emerging roles for E2F: Beyond the G1/S transition and DNA replication", *Cancer Cell,* vol. 3, no. 4, pp. 311-316.

Camarero, N., Mascaró, C., Mayordomo, C., Vilardell, F., Haro, D. and Marrero, P.F. (2006) "Ketogenic HMGCS2 is a c-Myc target gene expressed in

differentiated cells of human colonic epithelium and down-regulated in colon cancer", *Molecular Cancer Research,* vol. 4, no. 9, pp. 645-653.

Campanero, M.R., Armstrong, M.I. and Flemington, E.K. (2000) "CpG methylation as a mechanism for the regulation of E2F activity", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 97, no. 12, pp. 6481-6486.

Carcagno, A.L., Marazita, M.C., Ogara, M.F., Ceruti, J.M., Sonzogni, S.V., Scassa, M.E., Giono, L.E. and Cánepa, E.T. (2011) "E2F1-mediated upregulation of p19INK4d determines its periodic expression during cell cycle and regulates cellular proliferation", *PLoS ONE,* vol. 6, no. 7, art no. e241938.

Carcagno, A.L., Ogara, M.F., Sonzogni, S.V., Marazita, M.C., Sirkin, P.F., Ceruti, J.M. and Cánepa, E.T. (2009) "E2F1 transcription is induced by genotoxic stress through ATM/ATR activation", *IUBMB life,* vol. 61, no. 5, pp. 537-543.

Carteau, S., Hoffmann, C. and Bushman, F. (1998) "Chromosome structure and human immunodeficiency virus type 1 cDNA integration: Centromeric alphoid repeats are a disfavored target", *Journal of virology,* vol. 72, no. 5, pp. 4005-4014.

Cassani, B., Montini, E., Maruggi, G., Ambrosi, A., Mirolo, M., Selleri, S., Biral, E., Frugnoli, I., Hernandez-Trujillo, V., Di Serio, C., Roncarolo, M.G., Naldini, L., Mavilio, F. and Aiuti, A. (2009) "Integration of retroviral vectors induces minor changes in the transcriptional activity of T cells from ADA-SCID patients treated with gene therapy", *Blood,* vol. 114, no. 17, pp. 3546-3556.

Cattoglio, C., Facchini, G., Sartori, D., Antonelli, A., Miccio, A., Cassani, B., Schmidt, M., Von Kalle, C., Howe, S., Thrasher, A.J., Aiuti, A., Ferrari, G., Recchia, A. and Mavilio, F. (2007) "Hot spots of retroviral integration in human CD34+ hematopoietic cells", *Blood,* vol. 110, no. 6, pp. 1770-1778.

Cavazzana-Calvo, M., Hacein-Bey, S., De Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L., Bousso, P., Deist, F.L., Fischer, A. and Kohn, D.B. (2000) "Gene therapy for XSCID: The first success of gene therapy", *Pediatric research,* vol. 48, no. 5, pp. 578.

Cavazzana-Calvo, M. and Hacein-Bey-Abina, S. (2001) "Correction of genetic blood defects by gene transfer", *Current opinion in hematology,* vol. 8, no. 6, pp. 360-367.

Cavazzana-Calvo, M., Lagresle, C., Hacein-Bey-Abina, S. and Fischer, A. (2005) "Gene therapy for severe combined immunodeficiency", *Annual review of medicine,* vol. 56, pp. 585-602

Cavazzana-Calvo, M., Lagresle, C., Hacein-Bey-Abina, S. and Fischer, A. (2005) *Gene therapy for severe combined immunodeficiency*.

Cavazzana-Calvo, M., Payen, E., Negre, O., Wang, G., Hehir, K., Fusil, F., Down, J., Denaro, M., Brady, T., Westerman, K., Cavallesco, R., Gillet-Legrand, B., Caccavelli, L., Sgarra, R., Maouche-Chrétien, L., Bernaudin, F., Girot, R., Dorazio, R., Mulder, G.-., Polack, A., Bank, A., Soulier, J., Larghero, J., Kabbara, N., Dalle, B., Gourmel, B., Socie, G., Chrétien, S., Cartier, N., Aubourg, P., Fischer, A., Cornetta, K., Galacteros, F., Beuzard, Y., Gluckman, E., Bushman, F., Hacein-Bey-Abina, S. and Leboulch, P. (2010) "Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia", *Nature,* vol. 467, no. 7313, pp. 318-322.

Ceccarelli, C., Santini, D., Chieco, P., Taffurelli, M., Gamberini, M., Pileri, S.A. and Marrano, D. (1998) "Retinoblastoma (RB1) gene product expression in breast carcinoma. Correlation with Ki-67 growth fraction and biopathological profile", *Journal of clinical pathology,* vol. 51, no. 11, pp. 818-824.

Ceci, J.D., Patriotis, C.P., Tsatsanis, C., Makris, A.M., Kovatch, R., Swing, D.A., Jenkins, N.A., Tsichlis, P.N. and Copeland, N.G. (1997) "Tpl-2 is an oncogenic kinase that is activated by carboxy-terminal truncation", *Genes and Development,* vol. 11, no. 6, pp. 688-700.

Challita, P.-. and Kohn, D.B. (1994) "Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 91, no. 7, pp. 2567-2571.

Challita, P.-., Skelton, D., El-Khoueiry, A., Yu, X.-., Weinberg, K. and Kohn, D.B. (1995) "Multiple modifications in cis elements of the long terminal repeat of retroviral vectors lead to increased expression and decreased DNA methylation in embryonic carcinoma cells", *Journal of virology,* vol. 69, no. 2, pp. 748-755.

Chan, B.C.-., Ching, A.K.-., To, K.-., Leung, J.C.-., Chen, S., Li, Q., Lai, P.B.-., Tang, N.L.-., Shaw, P.-., Chan, J.Y.-., James, A.E., Lai, K.-., Lim, P.-., Lee, K.K.-. and Chui, Y.-. (2008) "BRE is an antiapoptotic protein in vivo and overexpressed in human hepatocellular carcinoma", *Oncogene,* vol. 27, no. 9, pp. 1208-1217.

Chan, J.Y.-., Li, L., Miao, J., Cai, D.-., Lee, K.K.-. and Chui, Y.-. (2010) "Differential expression of a novel gene BRE (TNFRSF1A modulator/BRCC45) in response to stress and biological signals", *Molecular biology reports,* vol. 37, no. 1, pp. 363-368.

Chellappan, S., Kraus, V.B., Kroger, B., Munger, K., Howley, P.M., Phelps, W.C. and Nevins, J.R. (1992) "Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 89, no. 10, pp. 4549-4553.

Chellappan, S.P., Hiebert, S., Mudryj, M., Horowitz, J.M. and Nevins, J.R. (1991) "The E2F transcription factor is a cellular target for the RB protein", *Cell,* vol. 65, no. 6, pp. 1053-1061.

Chen, C.M.A., Smith, D.M., Peters, M.A., Samson, M.E.S., Zitz, J., Tabin, C.J. and Cepko, C.L. (1999) "Production and design of more effective avian replication-incompetent retroviral vectors", *Developmental biology,* vol. 214, no. 2, pp. 370-384.

Chen, H.-., Tsai, S.-. and Leone, G. (2009) "Emerging roles of E2Fs in cancer: An exit from cell cycle control", *Nature Reviews Cancer,* vol. 9, no. 11, pp. 785-797.

Chen, J., Odenike, O. and Rowley, J.D. (2010) "Leukaemogenesis: More than mutant genes", *Nature Reviews Cancer,* vol. 10, no. 1, pp. 23-36.

Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y. and Li, E. (2007) "Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells", *Nature genetics,* vol. 39, no. 3, pp. 391-396.

Chen, X., Cao, X., Dong, W., Xia, M., Luo, S., Fan, Q. and Xie, J. (2010) "Cystatin M expression is reduced in gastric carcinoma and is associated with promoter hypermethylation", *Biochemical and biophysical research communications,* vol. 391, no. 1, pp. 1070-1074.

Chen, Y.-., Wang, Q., Lin, S.-., Chang, C.D., Chung, J. and Ying, S.-. (2006) "Activin signaling and its role in regulation of cell proliferation, apoptosis, and carcinogenesis", *Experimental biology and medicine,* vol. 231, no. 5, pp. 534-544.

Cherepanov, P. (2007) "LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro", *Nucleic acids research,* vol. 35, no. 1, pp. 113-124.

Choi, I.-. and Wu, T.-. (2005) "Epigenetic alterations in gastric carcinogenesis", *Cell research,* vol. 15, no. 4, pp. 247-254.

Chong, H., Starkey, W. and Vile, R.G. (1998) "A replication-competent retrovirus arising from a split-function packaging cell line was generated by recombination events between the vector, one of the packaging constructs, and endogenous retroviral sequences", *Journal of virology,* vol. 72, no. 4, pp. 2663-2670.

ChoraŻy, M., Bendich, A., Borenfreund, E., Ittensohn, O. L . and Hutchison D. J. (1963) "Uptake of mammalian chromosomes by mammalian cells", *Journal of cell biology* , no. 19, vol. 1, pp. 71–77.

CHORAZY, M., WILCZOK, T., JERGOVIC, D. and MENDECKI, J. (1964) "STUDIES ON ADSORPTION OF DEOXYRIBO- AND RIBONUCLEIC ACIDS BY NEOPLASTIC", *Acta - Unio Internationalis Contra Cancrum,* vol. 20, pp. 1003-1005.

Chui, Y.-., Ching, A.K.-., Chen, S., Yip, F.-., Rowlands, D.K., James, A.E., Lee, K.K.-. & Chan, J.Y.-. (2010) "BRE over-expression promotes growth of hepatocellular carcinoma", *Biochemical and biophysical research communications,* vol. 391, no. 3, pp. 1522-1525.

Chun, N. & Park, M. (2012) "The impact of clinical trial insurance coverage mandates on racial minorities and low income individuals", *Applied Economics,* vol. 44, no. 15, pp. 1977-1984.

Ciceri, F., Bonini, C., Gallo-Stampino, C. & Bordignon, C. (2005) "Modulation of GvHD by suicide-gene transduced donor T lymphocytes: Clinical applications in mismatched transplantation", *Cytotherapy,* vol. 7, no. 2, pp. 144-149.

Ciuffi, A. & Bushman, F.D. (2006) "Retroviral DNA integration: HIV and the role of LEDGF/p75", *Trends in Genetics,* vol. 22, no. 7, pp. 388-395.

Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R. & Bushman, F. (2005) "A role for LEDGF/p75 in targeting HIV DNA integration", *Nature medicine,* vol. 11, no. 12, pp. 1287-1289.

Ciuffi, A., Mitchell, R.S., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R. & Bushman, F.D. (2006) "Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts", *Molecular Therapy,* vol. 13, no. 2, pp. 366-373.

Classon, M. & Harlow, E. (2002) "The retinoblastoma tumour suppressor in development and cancer", *Nature Reviews Cancer,* vol. 2, no. 12, pp. 910-917.

Clements, C.M., McNally, R.S., Conti, B.J., Mak, T.W. & Ting, J.P.-. (2006) "DJ-1, a cancer- and Parkinson's disease-associated protein, stabilizes the antioxidant transcriptional master regulator Nrf2", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 103, no. 41, pp. 15091-15096.

Clements, J.E. & Christine Zink, M. (1996) "Molecular biology and pathogenesis of animal lentivirus infections", *Clinical microbiology reviews,* vol. 9, no. 1, pp. 100-117.

Coffin, J.M. (1996) "Retrovirus restriction revealed", *Nature,* vol. 382, no. 6594, pp. 762-763.

Coffin, J.M. (1979) "Structure, replication, and recombination of retrovirus genomes: Some unifying hypotheses", *Journal of General Virology,* vol. 42, no. 1, pp. 1-26.

Condiotti, R., Curran, M.A., Nolan, G.P., Giladi, H., Ketzinel-Gilad, M., Gross, E. and Galun, E. (2004) "Prolonged liver-specific transgene expression by a non-primate lentiviral vector", *Biochemical and biophysical research communications,* vol. 320, no. 3, pp. 998-1006.

Conklin, K.F. & Groudine, M. (1986) "Varied interactions between proviruses and adjacent host chromatin.", *Molecular and cellular biology,* vol. 6, no. 11, pp. 3999-3407.

Conley, A.B., Piriyapongsa, J. & Jordan, I.K. (2008) "Retroviral promoters in the human genome", *Bioinformatics,* vol. 24, no. 14, pp. 1563-1567.

Cooley, L., Kelley, R. & Spradling, A. (1988) "Insertional mutagenesis of the Drosophila genome with single P elements", *Science,* vol. 239, no. 4844, pp. 1121-1128.

Copeland, N.G., Jenkins, N.A. & Lee, B.K. (1983a) "Association of the lethal yellow (A(y)) coat color mutation with an ecotropic murine leukemia virus genome", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 80, no. 1 I, pp. 247-249.

Copeland, T.D., Oroszlan, S. & Kalyanaraman, V.S. (1983b) "Complete amino acid sequence of human T-cell leukemia virus structural protein p15", *FEBS letters,* vol. 162, no. 2, pp. 390-395.

Cosset, F.-. & Russell, S.J. (1996) "Targeting retrovirus entry", *Gene therapy,* vol. 3, no. 11, pp. 946-956.

Coutelle, C., Themis, M., Waddington, S.N., Buckley, S.M.K., Gregory, L.G., Nivsarkar, M.S., David, A.L., Peebles, D., Weisz, B. & Rodeck, C. (2005) "Gene therapy progress and prospects: Fetal gene therapy - First proofs of concept - Some adverse effects", *Gene therapy,* vol. 12, no. 22, pp. 1601-1607.

Cronin, J., Zhang, X.-. & Reiser, J. (2005) "Altering the tropism of lentiviral vectors through pseudotyping", *Current Gene Therapy,* vol. 5, no. 4, pp. 387-398.

Culver, K.W., Osborne, W.R.A., Miller, A.D., Fleisher, T.A., Berger, M., Anderson, W.F. & Blaese, R.M. (1991) "Correction of ADA deficiency in human T lymphocytes using retroviral-mediated gene transfer", *Transplantation proceedings,* vol. 23, no. 1 I, pp. 170-171.

Cuozzo, C., Porcellini, A., Angrisano, T., Morano, A., Lee, B., Di Pardo, A., Messina, S., Iuliano, R., Fusco, A., Santillo, M.R., Muller, M.T., Chiariotti, L., Gottesman, M.E. & Avvedimento, E.V. (2007) "DNA damage, homology-directed repair, and DNA methylation", *PLoS Genetics,* vol. 3, no. 7, pp. 1144-1162.

Daly, G. & Chernajovsky, Y. (2000) "Recent developments in retroviral-mediated gene transduction", *Molecular Therapy,* vol. 2, no. 5, pp. 423-434.

Damico, R. & Bates, P. (2000) "Soluble receptor-induced retroviral infection of receptor-deficient cells", *Journal of virology,* vol. 74, no. 14, pp. 6469-6475.

Das, P.M. & Singal, R. (2004) "DNA methylation and cancer", *Journal of Clinical Oncology,* vol. 22, no. 22, pp. 4632-4642.

Davé, U.P., Akagi, K., Tripathi, R., Cleveland, S.M., Thompson, M.A., Yi, M., Stephens, R., Downing, J.R., Jenkins, N.A. & Copeland, N.G. (2009) "Murine leukemias with retroviral insertions at Lmo2 are predictive of the leukemias induced in SCID-X1 patients following retroviral gene therapy", *PLoS Genetics,* vol. 5, no. 5. art no. e1000491.

Davé, U.P., Jenkins, N.A. & Copeland, N.G. (2004) "Gene Therapy Insertional Mutagenesis Insights", *Science,* vol. 303, no. 5656, pp. 333.

David, A.L. & Peebles, D. (2008) "Gene therapy for the fetus: is there a future?", *Best Practice and Research in Clinical Obstetrics and Gynaecology,* vol. 22, no. 1, pp. 203-218.

David, G., Grandinetti, K.B., Finnerty, P.M., Simpson, N., Chu, G.C. & DePinho, R.A. (2008) "Specific requirement of the chromatin modifier mSin3B in cell cycle exit and cellular differentiation", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 105, no. 11, pp. 4168-4172.

De Palma, M., Montini, E., Santoni De Sio, F.R., Benedicenti, F., Gentile, A., Medico, E. & Naldini, L. (2005) "Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoietic cells", *Blood,* vol. 105, no. 6, pp. 2307-2315.

Degregori, J., Leone, G., Miron, A., Jakoi, L. & Nevins, J.R. (1997) "Distinct roles for E2F proteins in cell growth control and apoptosis", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 94, no. 14, pp. 7245-7250.

Deichmann, A., Hacein-Bey-Abina, S., Schmidt, M., Garrigue, A., Brugman, M.H., Hu, J., Glimm, H., Gyapay, G., Prum, B., Fraser, C.C., Fischer, N., Schwarzwaelder, K., Siegler, M.-., De Ridder, D., Pike-Overzet, K., Howe, S.J., Thrasher, A.J., Wagemaker, G., Abel, U., Staal, F.J.T., Delabesse, E., Villeval, J.-., Aronow, B., Hue, C., Prinz, C., Wissler, M., Klanke, C., Weissenbach, J., Alexander, I., Fischer, A., Von Kalle, C. & Cavazzana-Calvo, M. (2007) "Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy", *Journal of Clinical Investigation,* vol. 117, no. 8, pp. 2225-2232.

Delassus, S., Cheynier, R. & Wain-Hobson, S. (1991) "Evolution of human immunodeficiency virus type 1 nef and long terminal repeat sequences over 4 years in vivo and in vitro", *Journal of virology,* vol. 65, no. 1, pp. 225-231.

Delenda, C. (2004) "Lentiviral vectors: Optimization of packaging, transduction and gene expression", *Journal of Gene Medicine,* vol. 6, no. SUPPL. 1, pp. S125-S138.

Denchi, E.L. & Helin, K. (2005) "E2F1 is crucial for E2F-dependent apoptosis", *EMBO reports,* vol. 6, no. 7, pp. 661-667.

Deng, J., Shoemaker, R., Xie, B., Gore, A., Leproust, E.M., Antosiewicz-Bourget, J., Egli, D., Maherali, N., Park, I.-., Yu, J., Daley, G.Q., Eggan, K., Hochedlinger, K., Thomson, J., Wang, W., Gao, Y. & Zhang, K. (2009) "Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming", *Nature biotechnology,* vol. 27, no. 4, pp. 353-360.

Devaskar, S.U. & Raychaudhuri, S. (2007) "Epigenetics - A science of heritable biological adaptation", *Pediatric research,* vol. 61, no. 5 PART 2 SUPPL., pp. 1R-4R.

Dieudonné, M., Maiuri, P., Biancotto, C., Knezevich, A., Kula, A., Lusic, M. & Marcello, A. (2009) "Transcriptional competence of the integrated HIV-1 provirus at the nuclear periphery", *EMBO Journal,* vol. 28, no. 15, pp. 2231-2243.

Dinauer, M.C., Bjorgvinsdottir, H., Ding, C., Pech, N., Morgenstern, D. & Gifford, M. (1996) "Retroviral-mediated gene transfer of gp91phox rescues the defect in host defense against aspergillus fumigatus in mice with x-linked chronic granulomatous disease", *Journal of Investigative Medicine,* vol. 44, no. 3.

Dinauer, M.C., Li, L.L., Björgvinsdóttir, H., Ding, C. & Pech, N. (1999) "Long-term correction of phagocyte NADPH oxidase activity retroviral- mediated gene transfer in murine X-linked chronic granulomatous disease", *Blood,* vol. 94, no. 3, pp. 914-922.

Dodge, J.E., Ramsahoye, B.H., Wo, Z.G., Okano, M. & Li, E. (2002) "De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation", *Gene,* vol. 289, no. 1-2, pp. 41-48.

Dodget, J.E., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N. & Li, E. (2005) "Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization", *Journal of Biological Chemistry,* vol. 280, no. 18, pp. 17986-17991.

Doerfler, W., Remus, R., Müller, K., Heller, H., Hohlweg, U. & Schubbert, R. (2001) "The fate of foreign DNA in mammalian cells and organisms.", *Developments in biologicals,* vol. 106, pp. 89-97; discussion 143.

Doerfler, W., Schubbert, R., Heller, H., Kämmer, C., Hilger-Eversheim, K., Knoblauch, M. & Remus, R. (1997) "Integration of foreign DNA and its consequences in mammalian systems", *Trends in biotechnology,* vol. 15, no. 8, pp. 297-301.

Dolinoy, D.C., Weidman, J.R. & Jirtle, R.L. (2007) "Epigenetic gene regulation: Linking early developmental environment to adult disease", *Reproductive Toxicology,* vol. 23, no. 3, pp. 297-307.

Delgado, D.M. & León, J. (2006) "Gene expression regulation and cancer", *Clinical and Translational Oncology,* vol. 8, no. 11, pp. 780-787.

Du, Y., Jenkins, N.A. & Copeland, N.G. (2005) "Insertional mutagenesis identifies genes that promote the immortalization of primary bone marrow progenitor cells", *Blood,* vol. 106, no. 12, pp. 3932-3939.

Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D. & Naldini, L. (1998) "A third-generation lentivirus vector with a conditional packaging system", *Journal of virology,* vol. 72, no. 11, pp. 8463-8471.

Dunbar, C.E. & Larochelle, A. (2010) "Gene therapy activates EVI1, destabilizes chromosomes", *Nature medicine,* vol. 16, no. 2, pp. 163-165.

Ehrlich, M. (2006) "Cancer-linked DNA hypomethylation and its relationship to hypermethylation", *Current topics in microbiology and immunology*, vol. 310, pp. 251-274.

Ehrlich, M. (2002) "DNA hypomethylation, cancer, the immunodeficiency, centromeric region instability, facial anomalies syndrome and chromosomal rearrangements", *Journal of Nutrition,* vol. 132, no. 8 SUPPL., pp. 2424S-2429S.

Ehrlich, M. (2002) "DNA methylation in cancer: Too much, but also too little", *Oncogene,* vol. 21, no. 35 REV. ISS. 3, pp. 5400-5413.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.-., Midgett, R.M., Kuo, K.C., Mccune, R.A. & Gehrke, C. (1982) "Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells", *Nucleic acids research,* vol. 10, no. 8, pp. 2709-2721.

Ehrlich, M., Jackson, K. & Weemaes, C. (2006) "Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF)", *Orphanet Journal of Rare Diseases,* vol. 1, no. 1.

Ellingsen, S., Laplante, M.A., König, M., Kikuta, H., Furmanek, T., Hoivik, E.A. & Becker, T.S. (2005) "Large-scale enhancer detection in the zebrafish genome", *Development,* vol. 132, no. 17, pp. 3799-3811.

Ellison, V., Abrams, H., Roe, T., Lifson, J. & Brown, P. (1990) "Human immunodeficiency virus integration in a cell-free system", *Journal of virology,* vol. 64, no. 6, pp. 2711-2715.

El-Osta, A. (2003) "DNMT cooperativity - The developing links between methylation, chromatin structure and cancer", *BioEssays,* vol. 25, no. 11, pp. 1071-1084.

Emerman, M. (2006) "How TRIM5α defends against retroviral invasions", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 103, no. 14, pp. 5249-5250.

Emery, D.W., Yannaki, E., Tubb, J. & Stamatoyannopoulos, G. (2000) "A chromatin insulator protects retrovirus vectors from chromosomal position effects",

Proceedings of the National Academy of Sciences of the United States of America, vol. 97, no. 16, pp. 9150-9155.

Esteller, M. (2008) "Molecular origins of cancer: Epigenetics in cancer", *New England Journal of Medicine,* vol. 358, no. 11, pp. 1148-1159+1096.

Fackenthal, J.D. & Godley, L.A. (2008) "Aberrant RNA splicing and its functional consequences in cancer cells", *DMM Disease Models and Mechanisms,* vol. 1, no. 1, pp. 37-42.

Fang, J.-., Mikovits, J.A., Bagni, R., Petrow-Sadowski, C.L. & Ruscetti, F.W. (2001) "Infection of lymphoid cells by integration-defective human immunodeficiency virus type 1 increases de novo methylation", *Journal of virology,* vol. 75, no. 20, pp. 9753-9761.

Fehse, B., Kustikova, O.S., Bubenheim, M. & Baum, C. (2004) "Pois(s)on - It's a question of dose..", *Gene therapy,* vol. 11, no. 11, pp. 879-881.

Fehse, B., Kustikova, O.S., Li, Z., Wahlers, A., Bohn, W., Beyer, W.R., Chalmers, D., Tiberghien, P., Kühlcke, K., Zander, A.R. & Baum, C. (2002) "A novel 'sort-suicide' fusion gene vector for T cell manipulation", *Gene therapy,* vol. 9, no. 23, pp. 1633-1638.

Feng, W., Shen, L., Wen, S., Rosen, D.G., Jelinek, J., Hu, X., Huan, S., Huang, M., Liu, J., Sahin, A.A., Hunt, K.K., Bast Jr., R.C., Shen, Y., Issa, J.-.J. & Yu, Y. (2007) "Correlation between CpG methylation profiles and hormone receptor status in breast cancers", *Breast Cancer Research,* vol. 9, no. 4.

Field, S.J., Tsai, F.-., Kuo, F., Zubiaga, A.M., Kaelin Jr., W.G., Livingston, D.M., Orkin, S.H. & Greenberg, M.E. (1996) "E2F-1 Functions in mice to promote apoptosis and suppress proliferation", *Cell,* vol. 85, no. 4, pp. 549-561.

Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick, K., Margolick, J., Brookmeyer, R., Gallant, J., Markowitz, M., Ho, D.D., Richman, D.D. & Siliciano, R.F. (1997) "Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy", *Science,* vol. 278, no. 5341, pp. 1295-1300.

Fischer, A., Hacein-Bey-Abina, S. & Cavazzana-Calvo, M. (2010) "20 years of gene therapy for SCID", *Nature immunology,* vol. 11, no. 6, pp. 457-460.

Flanagan, J.M. (2007) "Host epigenetic modifications by oncogenic viruses", *British journal of cancer,* vol. 96, no. 2, pp. 183-188.

Frazer, K.A., Tao, H., Osoegawa, K., de Jong, P.J., Chen, X., Doherty, M.F. & Cox, D.R. (2004) "Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional", *Genome research,* vol. 14, no. 3, pp. 367-372.

Friedmann, T. (1992) "A brief history of gene therapy", *Nature genetics,* vol. 2, no. 2, pp. 93-98.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. & Paul, C.L. (1992) "A genomic sequencing protocol that yields a positive display of 5- methylcytosine residues in individual DNA strands", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 89, no. 5, pp. 1827-1831.

Fuss, B., Becker, T., Zinke, I. & Hoch, M. (2006) "The cytohesin Steppke is essential for insulin signalling in Drosophila", *Nature,* vol. 444, no. 7121, pp. 945-948.

Ganser-Pornillos, B.K., Chandrasekaran, V., Pornillos, O., Sodroski, J.G., Sundquist, W.I. & Yeager, M. (2011) "Hexagonal assembly of a restricting TRIM5alpha protein.", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 108, no. 2, pp. 534-539.

Gaspar, H.B., Cooray, S., Gilmour, K.C., Parsley, K.L., Adams, S., Howe, S.J., Al Ghonaium, A., Bayford, J., Brown, L., Davies, E.G., Kinnon, C. and Thrasher, A.J. (2011) "Immunodeficiency: Long-term persistence of a polyclonal t cell repertoire after gene therapy for X-linked severe combined immunodeficiency", *Science Translational Medicine,* vol. 3, no. 97.

Gaspar, H.B., Parsley, K.L., Howe, S., King, D., Gilmour, K.C., Sinclair, J., Brouns, G., Schmidt, M., Kalle, C.V., Barington, T., Jakobsen, M.A., Christensen, H.O., Ghonaium, A.A., White, H.N., Smith, J.L., Levinsky, P.R.J., Ali, P.R.R., Kinnon, P.C. & Thrasher, P.A.J. (2004) "Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector", *Lancet,* vol. 364, no. 9452, pp. 2181-2187.

Ginn, S.L., Liao, S.H., Dane, A.P., Hu, M., Hyman, J., Finnie, J.W., Zheng, M., Cavazzana-Calvo, M., Alexander, S.I., Thrasher, A.J. & Alexander, I.E. (2010) "Lymphomagenesis in SCID-X1 mice following lentivirus-mediated phenotype correction independent of insertional mutagenesis and γc overexpression", *Molecular Therapy,* vol. 18, no. 5, pp. 965-976.

Girard, L., Hanna, Z., Beaulieu, N., Hoemann, C.D., Simard, C., Kozak, C.A. & Jolicoeur, P. (1996) "Frequent provirus insertional mutagenesis of Notch1 in thymomas of MMTV(D)/myc transgenic mice suggests a collaboration of c-myc and Notch1 for oncogenesis", *Genes and Development,* vol. 10, no. 15, pp. 1930-1944.

Goff, S.P. (1987) Gene isolation by retroviral tagging. *Methods in enzymology*, vol 152, no. C, pp. 459-481.

Goi, T., Kawasaki, M., Yamazaki, T., Koneri, K., Katayama, K., Hirose, K. & Yamaguchi, A. (2003) "Ascending colon cancer with hepatic metastasis and cholecystolithiasis in a patient with situs inversus totalis without any expression of uvrag mrna: Report of a case", *Surgery today,* vol. 33, no. 9, pp. 702-706.

Gonda, T.J., Cory, S., Sobieszczuk, P., Holtzman, D. & Adams, J.M. (1987) "Generation of altered transcripts by retroviral insertion within the c-myb gene in two murine monocytic leukemias", *Journal of virology,* vol. 61, no. 9, pp. 2754-2763.

Gonda, T.J., Sheiness, D.K. & Bishop, J.M. (1982) "Transcripts from the cellular homologs of retroviral oncogenes: Distribution among chicken tissues", *Molecular and cellular biology,* vol. 2, no. 6, pp. 617-624.

Gregory, L.G., Waddington, S.N., Holder, M.V., Mitrophanous, K.A., Buckley, S.M.K., Mosley, K.L., Bigger, B.W., Ellard, F.M., Walmsley, L.E., Lawrence, L., Al-Allaf, F., Kingsman, S., Coutelle, C. & Themis, M. (2004) "Highly efficient EIAV-mediated in utero gene transfer and expression in the major muscle groups affected by Duchenne muscular dystrophy", *Gene therapy,* vol. 11, no. 14, pp. 1117-1125.

Hacein-Bey-Abina, S., Fischer, A. & Cavazzana-Calvo, M. (2002) "Gene therapy of X-linked severe combined immunodeficiency", *International journal of hematology,* vol. 76, no. 4, pp. 295-298.

Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., Asnafi, V., MacIntyre, E., Dal Cortivo, L., Radford, I., Brousse, N., Sigaux, F., Moshous, D., Hauer, J., Borkhardt, A., Belohradsky, B.H., Wintergerst, U., Velez, M.C., Leiva, L., Sorensen, R., Wulffraat, N., Blanche, S., Bushman, F.D., Fischer, A. & Cavazzana-Calvo, M. (2008) "Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1", *Journal of Clinical Investigation,* vol. 118, no. 9, pp. 3132-3142.

Hacein-Bey-Abina, S., Hauer, J., Lim, A., Picard, C., Wang, G.P., Berry, C.C., Martinache, C., Rieux-Laucat, F., Latour, S., Belohradsky, B.H., Leiva, L., Sorensen, R., Debré, M., Casanova, J.L., Blanche, S., Durandy, A., Bushman, F.D., Fischer, A. & Cavazzana-Calvo, M. (2010) "Efficacy of gene therapy for X-linked severe combined immunodeficiency", *New England Journal of Medicine,* vol. 363, no. 4, pp. 355-364.

Hacein-Bey-Abina, S., Le Deist, F., Carlier, F., Bouneaud, C., Hue, C., De Villartay, J.-., Thrasher, A.J., Wulffraat, N., Sorensen, R., Dupuis-Girod, S., Fischer, A., Graham Davies, E., Kuis, W., Leiva, L. & Cavazzana-Calvo, M. (2002) "Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy", *New England Journal of Medicine,* vol. 346, no. 16, pp. 1185-1193.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E., Sorensen, R., Forster, A., Fraser, P., Cohen, J.I., De Saint Basile, G., Alexander, I., Wintergerst, U., Frebourg, T., Aurias, A., Stoppa-Lyonnet, D., Romana, S., Radford-Weiss, I., Gross, F., Valensi, F., Delabesse, E., Macintyre, E., Sigaux, F., Soulier, J., Leiva, L.E., Wissler, M., Prinz, C., Rabbitts, T.H., Le Deist, F., Fischer, A. & Cavazzana-Calvo, M. (2003) "LMO2-Associated Clonal T Cell

Proliferation in Two Patients after Gene Therapy for SCID-X1", *Science,* vol. 302, no. 5644, pp. 415-419.

Hacker, C.V., Vink, C.A., Wardell, T.W., Lee, S., Treasure, P., Kingsman, S.M., Mitrophanous, K.A. & Miskin, J.E. (2006) "The integration profile of EIAV-based vectors", *Molecular Therapy,* vol. 14, no. 4, pp. 536-545.

Haffter, P., Granato, M., Brand, M., Mullins, M.C., Hammerschmidt, M., Kane, D.A., Odenthal, J., Van Eeden, F.J.M., Jiang, Y.-., Heisenberg, C.-., Kelsh, R.N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C. & Nüsslein-Volhard, C. (1996) "The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio", *Development,* vol. 123, pp. 1-36.

Hafner, M., Schmitz, A., Grüne, I., Srivatsan, S.G., Paul, B., Kolanus, W., Quast, T., Kremmer, E., Bauer, I. & Famulok, M. (2006) "Inhibition of cytohesins by SecinH3 leads to hepatic insulin resistance", *Nature,* vol. 444, no. 7121, pp. 941-944.

Hanlon, L., Barr, N.I., Blyth, K., Stewart, M., Haviernik, P., Wolff, L., Weston, K., Cameron, E.R. & Neil, J.C. (2003) "Long-range effects of retroviral insertion on c-myb: Overexpression may be obscured by silencing during tumor growth in vitro", *Journal of virology,* vol. 77, no. 2, pp. 1059-1068.

Hantzopoulos, P.A., Suri, C., Glass, D.J., Goldfarb, M.P. & Yancopoulos, G.D. (1994) "The low affinity NGF receptor, p75, can collaborate with each of the Trks to potentiate functional responses to the neurotrophins", *Neuron,* vol. 13, no. 1, pp. 187-201.

Harbers, K., Kuehn, M., Delius, H. & Jaenisch, R. (1984) "Insertion of retrovirus into the first intron of alpha 1(I) collagen gene to embryonic lethal mutation in mice.", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 81, no. 5, pp. 1504-1508.

Harbers, K., Schnieke, A. & Stuhlmann, H. (1981) "DNA methylation and gene expression: Endogenous retroviral genome becomes infectious after molecular cloning", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 78, no. 12 II, pp. 7609-7613.

Hashimoto, S., Onodera, Y., Hashimoto, A., Tanaka, M., Hamaguchi, M., Yamada, A. & Sabe, H. (2004) "Requirement for Arf6 in breast cancer invasive activities", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 101, no. 17, pp. 6647-6652.

Hauswirth, W.W., Aleman, T.S., Kaushal, S., Cideciyan, A.V., Schwartz, S.B., Wang, L., Conlon, T.J., Boye, S.L., Flotte, T.R., Byrne, B.J. & Jacobson, S.G. (2008) "Treatment of Leber congenital amaurosis due to RPE65 mutations by ocular subretinal injection of adeno-associated virus gene vector: Short-term results of a phase I trial", *Human Gene Therapy,* vol. 19, no. 10, pp. 979-990.

Hayakawa, J., Washington, K., Uchida, N., Phang, O., Kang, E.M., Hsieh, M.M. & Tisdale, J.F. (2009) "Long-term vector integration site analysis following retroviral mediated gene transfer to hematopoietic stem cells for the treatment of HIV infection", *PLoS ONE,* vol. 4, no. 1, art no. e4211.

Hejnar, J., Elleder, D., Hájková, P., Walter, J., Blažková, J. & Svoboda, J. (2003) "Demethylation of host-cell DNA at the site of avian retrovirus integration", *Biochemical and biophysical research communications,* vol. 311, no. 3, pp. 641-648.

Heller, H., Kämmer, C., Wilgenbus, P. & Doerfler, W. (1995) "Chromosomal insertion of foreign (adenovirus type 12, plasmid, or bacteriophage λ) DNA is associated with enhanced methylation of cellular DNA segments", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 92, no. 12, pp. 5515-5519.

Hendrie, P.C., Huo, Y., Stolitenko, R.B. & Russell, D.W. (2008) "A rapid and quantitative assay for measuring neighboring gene activation by vector proviruses", *Molecular Therapy,* vol. 16, no. 3, pp. 534-540.

Herman, S.A. & Coffin, J.M. (1987) "Efficient packaging of readthrough RNA in ALV: Implications for oncogene transduction", *Science,* vol. 236, no. 4803, pp. 845-848.

Herman, S.A. & Coffin, J.M. (1986) "Differential transcription from the long terminal repeats of integrated avian leukosis virus DNA", *Journal of virology,* vol. 60, no. 2, pp. 497-505.

Hernandez, L.D., Peters, R.J., Delos, S.E., Young, J.A.T., Agard, D.A. & White, J.M. (1997) "Activation of a retroviral membrane fusion protein: Soluble receptor- induced liposome binding of the ALSV envelope glycoprotein", *Journal of Cell Biology,* vol. 139, no. 6, pp. 1455-1464.

Higashimoto, T., Urbinati, F., Perumbeti, A., Jiang, G., Zarzuela, A., Chang, L.-., Kohn, D.B. & Malik, P. (2007) "The woodchuck hepatitis virus post-transcriptional regulatory element reduces readthrough transcription from retroviral vectors", *Gene therapy,* vol. 14, no. 17, pp. 1298-1304.

Holman, A.G. & Coffin, J.M. (2005) "Symmetrical base preferences surrounding HIV-1 and avian sarcoma/leukosis virus but not murine leukemia virus integration sites", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, no. 17, pp. 6103-6107.

Howe, S.J., Mansour, M.R., Schwarzwaelder, K., Bartholomae, C., Hubank, M., Kempski, H., Brugman, M.H., Pike-Overzet, K., Chatters, S.J., De Ridder, D., Gilmour, K.C., Adams, S., Thornhill, S.I., Parsley, K.L., Staal, F.J.T., Gale, R.E., Linch, D.C., Bayford, J., Brown, L., Quaye, M., Kinnon, C., Ancliff, P., Webb, D.K., Schmidt, M., Von Kalle, C., Gaspar, H.B. & Thrasher, A.J. (2008) "Insertional mutagenesis combined with acquired somatic mutations causes

leukemogenesis following gene therapy of SCID-X1 patients", *Journal of Clinical Investigation,* vol. 118, no. 9, pp. 3143-3150.

Hsieh, J.-., Yap, D., O'Connor, D.J., Fogal, V., Fallis, L., Chan, F., Zhong, S. & Lu, X. (2002) "Novel function of the cyclin A binding site of E2F in regulating p53-induced apoptosis in response to DNA damage", *Molecular and cellular biology,* vol. 22, no. 1, pp. 78-93.

Hu, J., Renaud, G., Golmes, T., Ferris, A., Hendrie, P.C., Donahue, R.E., Hughes, S.H., Wolfsberg, T.G., Russell, D.W. & Dunbar, C.E. (2008) "Reduced genotoxicity of avian sarcoma leukosis virus vectors in rhesus long-term repopulating cells compared to standard murine retrovirus vectors", *Molecular Therapy,* vol. 16, no. 9, pp. 1617-1623.

Hu, W.-. & Pathak, V.K. (2000) "Design of retroviral vectors and helper cells for gene therapy", *Pharmacological reviews,* vol. 52, no. 4, pp. 493-511.

Huang, Y., Ishiko, T., Nakada, S., Utsugisawa, T., Kato, T. & Yuan, Z.-. (1997) "Role for E2F in DNA damage-induced entry of cells into S phase", *Cancer research,* vol. 57, no. 17, pp. 3640-3643.

Huck-Hui, N. & Bird, A. (1999) "DNA methylation and chromatin modification", *Current Opinion in Genetics and Development,* vol. 9, no. 2, pp. 158-163.

Humeau, L.M., Binder, G.K., Lu, X., Slepushkin, V., Merling, R., Echeagaray, P., Pereira, M., Slepushkina, T., Barnett, S., Dropulic, L.K., Carroll, R., Levine, B.L., June, C.H. & Dropulic, B. (2004) "Efficient lentiviral vector-mediated control of HIV-1 replication in CD4 lymphocytes from diverse HIV+ infected patients grouped according to CD4 count and viral load", *Molecular Therapy,* vol. 9, no. 6, pp. 902-913.

Ikeda, Y., Collins, M.K.L., Radcliffe, P.A., Mitrophanous, K.A. & Takeuchi, Y. (2002) "Gene transduction efficiency in cells of different species by HIV and EIAV vectors", *Gene therapy,* vol. 9, no. 14, pp. 932-938.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R. & Bird, A.P. (2010) "Orphan CpG Islands Identify numerous conserved promoters in the mammalian genome", *PLoS Genetics,* vol. 6, no. 9, art no. e1001134.

Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., Humphray, S., Cox, T., Langford, C. and Bird, A. (2008) "A novel CpG island set identifies tissue-specific methylation at developmental gene loci", *PLoS Biology,* vol. 6, no. 1, pp. 0037-0051, art no. e22.

Jaenisch, R. & Bird, A. (2003) "Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals", *Nature genetics,* vol. 33, no. SUPPL., pp. 245-254.

Jahner, D. & Jaenisch, R. (1985) "Retrovirus-induced de novo methylation of flanking host sequences correlates with gene inactivity", *Nature,* vol. 315, no. 6020, pp. 594-597.

Jern, P. & Coffin, J.M. (2008) "Effects of retroviruses on host genome function", *Annual review of genetics*, vol 42, pp. 709-732.

Jiang, W., Kanter, M.R., Dunkel, I., Ramsay, R.G., Beemon, K.L. & Hayward, W.S. (1997) "Minimal truncation of the c-myb gene product in rapid-onset B-cell lymphoma?", *Journal of virology,* vol. 71, no. 9, pp. 6526-6533.

Jiang, Y.-., Bressler, J. & Beaudet, A.L. (2004) "Epigenetics and human disease", *Annual review of genomics and human genetics*, vol. 5, pp. 479-510.

Jimenez, D.F., Lee, C.I., O'Shea, C.E., Kohn, D.B. & Tarantal, A.F. (2005) "HIV-1-derived lentiviral vectors and fetal route of administration on transgene biodistribution and expression in rhesus monkeys", *Gene therapy,* vol. 12, no. 10, pp. 821-830.

Joachims, M.L., Marble, P.A., Laurent, A.B., Pastuszko, P., Paliotta, M., Blackburn, M.R. & Thompson, L.F. (2008) "Restoration of adenosine deaminase-deficient human thymocyte development in vitro by inhibition of deoxynucleoside kinases", *Journal of Immunology,* vol. 181, no. 11, pp. 8153-8161.

Johnson, W.E. (2007) "Assisted suicide for retroviruses", *Nature biotechnology,* vol. 25, no. 6, pp. 643-644.

Jones, P.A. & Baylin, S.B. (2002) "The fundamental role of epigenetic events in cancer", *Nature Reviews Genetics,* vol. 3, no. 6, pp. 415-428.

Jones, P.A. & Laird, P.W. (1999) "Cancer epigenetics comes of age", *Nature genetics,* vol. 21, no. 2, pp. 163-167.

Jonkers, J., Korswagen, H.C., Acton, D., Breuer, M. & Berns, A. (1997) "Activation of a novel proto-oncogene, Frat1, contributes to progression of mouse T-cell lymphomas", *EMBO Journal,* vol. 16, no. 3, pp. 441-450.

Kafri, T., Blömer, U., Peterson, D.A., Gage, F.H. & Verma, I.M. (1997) "Sustained expression of genes delivered directly into liver and muscle by lentiviral vectors", *Nature genetics,* vol. 17, no. 3, pp. 314-317.

Kanai, Y., Ushijima, S., Kondo, Y., Nakanishi, Y. & Hirohashi, S. (2001) "DNA methyltransferase expression and DNA methylation of CPG islands and peri-centromeric satellite regions in human colorectal and stomach cancers", *International Journal of Cancer,* vol. 91, no. 2, pp. 205-212.

Kang, Y., Moressi, C.J., Scheetz, T.E., Xie, L., Tran, D.T., Casavant, T.L., Ak, P., Benham, C.J., Davidson, B.L. & McCray Jr., P.B. (2006) "Integration site choice of a feline immunodeficiency virus vector", *Journal of virology,* vol. 80, no. 17, pp. 8820-8823.

Kang, Y., Xie, L., Tran, D.T., Stein, C.S., Hickey, M., Davidson, B.L. & McCray Jr., P.B. (2005) "Persistent expression of factor VIII in vivo following nonprimate lentiviral gene transfer", *Blood,* vol. 106, no. 5, pp. 1552-1558.

Karlin, S., Doerfler, W. & Cardon, L.R. (1994) "Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses?", *Journal of virology,* vol. 68, no. 5, pp. 2889-2897.

Katen, L.J., Januszeski, M.M., Anderson, W.F., Hasenkrug, K.J. & Evans, L.H. (2001) "Infectious entry by amphotropic as well as ecotropic murine leukemia viruses occurs through an endocytic pathway", *Journal of virology,* vol. 75, no. 11, pp. 5018-5026.

Kay, M.A., Glorioso, J.C. & Naldini, L. (2001) "Viral vectors for gene therapy: The art of turning infectious agents into vehicles of therapeutics", *Nature medicine,* vol. 7, no. 1, pp. 33-40.

Kennedy, D.R., Hartnett, B.J., Kennedy, J.S., Vernau, W., Moore, P.F., O'Malley, T., Burkly, L.C., Henthorn, P.S. and Felsburg, P.J. (2011) "Ex vivo γ-retroviral gene therapy of dogs with X-linked severe combined immunodeficiency and the development of a thymic T cell lymphoma", *Veterinary immunology and immunopathology,* vol. 142, no. 1-2, pp. 36-48.

Khulan, B., Thompson, R.F., Ye, K., Fazzari, M.J., Suzuki, M., Stasiek, E., Figueroa, M.E., Glass, J.L., Chen, Q., Montagna, C., Hatchwell, E., Selzer, R.R., Richmond, T.A., Green, R.D., Melnick, A. & Greally, J.M. (2006) "Comparative isoschizomer profiling of cytosine methylation: The HELP assay", *Genome research,* vol. 16, no. 8, pp. 1046-1055.

Kim, R.H., Peters, M., Jang, Y., Shi, W., Pintilie, M., Fletcher, G.C., DeLuca, C., Liepa, J., Zhou, L., Snow, B., Binari, R.C., Manoukian, A.S., Bray, M.R., Liu, F.-., Tsao, M.-. & Mak, T.W. (2005) "DJ-1, a novel regulator of the tumor suppressor PTEN", *Cancer Cell,* vol. 7, no. 3, pp. 263-273.

Kimura, N., Moribe, T., Iizuka, N., Miura, T., Tamatsukuri, S., Ishitsuka, H., Hamamoto, Y. & Oka, M. (2009) "Rapid and quantitative detection of CpG-methylation status using TaqMan PCR combined with methyl-binding-domain polypeptide", *Clinical biochemistry,* vol. 42, no. 10-11, pp. 1113-1122.

King, H.W.S. & Brookes, P. (1985) "On the mechanism of induction of resistance to 6-thioguanine in Chinese hamster V79 cells by 3-methylcholanthrene-diolepoxide", *Carcinogenesis,* vol. 6, no. 10, pp. 1471-1476.

King, W., Patel, M.D. & Lobel, L.I. (1985) "Insertion mutagenesis of embryonal carcinoma cells by retroviruses", *Science,* vol. 228, no. 4699, pp. 554-558.

Kingsman, S.M., Mitrophanous, K. & Olsen, J.C. (2005) "Potential oncogene activity of the woodchuck hepatitis post-transcriptional regulatory element (WPRE)", *Gene therapy,* vol. 12, no. 1, pp. 3-4.

Kitamura, Y., Ha Lee, Y.M. & Coffin, J.M. (1992) "Nonrandom integration of retroviral DNA in vitro: Effect of CpG methylation", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 89, no. 12, pp. 5532-5536.

Kizhatil, K. & Albritton, L.M. (1997) "Requirements for different components of the host cell cytoskeleton distinguish ecotropic murine leukemia virus entry via endocytosis from entry via surface fusion", *Journal of virology,* vol. 71, no. 10, pp. 7145-7156.

Knasmüller, S., Mersch-Sundermann, V., Kevekordes, S., Darroudi, F., Huber, W.W., Hoelzl, C., Bichler, J. and Majer, B.J. (2004) "Use of human-derived liver cell lines for the detection of environmental and dietary genotoxicants; Current state of knowledge", *Toxicology,* vol. 198, no. 1-3, pp. 315-328.

Kohn, D.B., Mitsuya, H., Ballow, M., Selegue, J.E., Barankiewicz, J., Cohen, A., Gelfand, E., Anderson, W.F. & Blaese, R.M. (1989) "Establishment and characterization of adenosine deaminase-deficient human T cell lines", *Journal of Immunology,* vol. 142, no. 11, pp. 3971-3977.

Kohn, D.B., Sadelain, M. & Glorioso, J.C. (2003) "Occurrence of leukaemia following gene therapy of X-linked SCID", *Nature Reviews Cancer,* vol. 3, no. 7, pp. 477-488.

Kolanus, W. (2007) "Guanine nucleotide exchange factors of the cytohesin family and their roles in signal transduction", *Immunological reviews,* vol. 218, no. 1, pp. 102-113.

Kool, J. and Berns, A. (2009) "High-throughput insertional mutagenesis screens in mice to identify oncogenic networks", *Nature Reviews Cancer,* vol. 9, no. 6, pp. 389-399.

Kovesdi, I., Reichel, R. & Nevins, J.R. (1986) "Identification of a cellular transcription factor involved in E1A trans-activation", *Cell,* vol. 45, no. 2, pp. 219-228.

Kung, H.-., Boerkoel, C. & Carter, T.H. (1991) "Retroviral mutagenesis of cellular oncogenes: A review with insights into the mechanisms of insertional activation", *Current topics in microbiology and immunology,* vol. 171, pp. 1-25.

Kuramoto, K., Follman, D., Hematti, P., Sellers, S., Laukkanen, M.O., Seggewiss, R., Metzger, M.E., Krouse, A., Donahue, R.E., Von Kalle, C. & Dunbar, C.E. (2004a) "The impact of low-dose busulfan on clonal dynamics in nonhuman primates", *Blood,* vol. 104, no. 5, pp. 1273-1280.

Kuramoto, K., Follmann, D.A., Hematti, P., Sellers, S., Agricola, B.A., Metzger, M.E., Donahue, R.E., Von Kalle, C. & Dunbar, C.E. (2004b) "Effect of chronic cytokine therapy on clonal dynamics in nonhuman primates", *Blood,* vol. 103, no. 11, pp. 4070-4077.

Kustikova, O., Brugman, M. & Baum, C. (2010) "The genomic risk of somatic gene therapy", *Seminars in cancer biology,* vol. 20, no. 4, pp. 269-278.

Kustikova, O.S., Geiger, H., Li, Z., Brugman, M.H., Chambers, S.M., Shaw, C.A., Pike-Overzet, K., De Ridder, D., Staal, F.J.T., Von Keudell, G., Cornils, K., Nattamai, K.J., Modlich, U., Wagemaker, G., Goodell, M.A., Fehse, B. & Baum, C. (2007) "Retroviral vector insertion sites associated with dominant hematopoietic clones mark "stemness" pathways", *Blood,* vol. 109, no. 5, pp. 1897-1907.

Kustikova, O.S., Modlich, U. & Fehse, B. (2009) "Retroviral insertion site analysis in dominant haematopoietic clones.", *Methods in molecular biology (Clifton, N.J.),* vol. 506, pp. 373-390.

Kustikova, O.S., Wahlers, A., Kühlcke, K., Stähle, B., Zander, A.R., Baum, C. & Fehse, B. (2003) "Dose finding with retroviral vectors: Correlation of retroviral vector copy numbers in single cells with gene transfer efficiency in a cell population", *Blood,* vol. 102, no. 12, pp. 3934-3937.

Kvinlaug, B.T., Chan, W.-., Bullinger, L., Ramaswami, M., Sears, C., Foster, D., Lazic, S.E., Okabe, R., Benner, A., Lee, B.H., De Silva, I., Valk, P.J.M., Delwel, R., Armstrong, S.A., Döhner, H., Gilliland, D.G. & Huntly, B.J.P. (2011) "Common and overlapping oncogenic pathways contribute to the evolution of acute myeloid leukemias", *Cancer research,* vol. 71, no. 12, pp. 4117-4129.

Kwong, P.D., Wyatt, R., Desjardins, E., Robinson, J., Culp, J.S., Hellmig, B.D., Sweet, R.W., Sodroski, J. & Hendrickson, W.A. (1999) "Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1)", *Journal of Biological Chemistry,* vol. 274, no. 7, pp. 4115-4123.

Lamikanra, A., Myers, K.A., Ferris, N., Mitrophanous, K.A. & Carroll, M.W. (2005) "In vivo evaluation of an EIAV vector for the systemic genetic delivery of therapeutic antibodies", *Gene therapy,* vol. 12, no. 12, pp. 988-998.

Lapeyre, J.N. & Becker, F.F. (1979) "5-Methylcytosine content of nuclear DNA during chemical hepatocarcinogenesis and in carcinomas which result", *Biochemical and biophysical research communications,* vol. 87, no. 3, pp. 698-705.

Lavie, L., Kitova, M., Maldener, E., Meese, E. & Mayer, J. (2005) "CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2)", *Journal of virology,* vol. 79, no. 2, pp. 876-883.

Lazarevich, N.L. & Alpern, D.V. (2008) "Hepatocyte nuclear factor 4 in epithelial development and carcinogenesis", *Molecular biology,* vol. 42, no. 5, pp. 699-709.

Lee, C., Chang, J.H., Lee, H.S. & Cho, Y. (2002) "Structural basis for the recognition of the E2F transactivation domain by the retinoblastoma tumor suppressor", *Genes and Development,* vol. 16, no. 24, pp. 3199-3212.

Lee, E.Y., Yuan, T.L., Danielian, P.S., West, J.C. & Lees, J.A. (2009) "E2F4 cooperates with pRB in the development of extra-embryonic tissues", *Developmental biology,* vol. 332, no. 1, pp. 104-115.

Lee, H.J., Kim, K.S., Kim, E.J., Choi, H.B., Lee, K.H., Park, I.H., Ko, Y., Jeong, S.W. & Kim, S.U. (2007) "Brain transplantation of immortalized human neural stem cells promotes functional recovery in mouse intracerebral hemorrhage stroke model", *Stem cells,* vol. 25, no. 5, pp. 1204-1212.

Lee, J.-. & Thorgeirsson, S.S. (2004) "Genome-scale profiling of gene expression in hepatocellular carcinoma: Classification, survival prediction, and identification of therapeutic targets", *Gastroenterology,* vol. 127, no. SUPPL., pp. S51-S55.

Lee, P.P., Fitzpatrick, D.R., Beard, C., Jessup, H.K., Lehar, S., Makar, K.W., Pérez-Melgosa, M., Sweetser, M.T., Schlissel, M.S., Nguyen, S., Cherry, S.R., Tsai, J.H., Tucker, S.M., Weaver, W.M., Kelso, A., Jaenisch, R. & Wilson, C.B. (2001) "A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival", *Immunity,* vol. 15, no. 5, pp. 763-774.

Lee, S., Lee, H.J., Kim, J.-., Lee, H.-., Jang, J.J. & Kang, G.H. (2003) "Aberrant CpG island hypermethylation along multistep hepatocarcinogenesis", *American Journal of Pathology,* vol. 163, no. 4, pp. 1371-1378.

Lei, H., Oh, S.P., Okano, M., Jüttermann, R., Goss, K.A., Jaenisch, R. & Li, E. (1996) "De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells", *Development,* vol. 122, no. 10, pp. 3195-3205.

Leroux, C., Cadoré, J.-. & Montelaro, R.C. (2004) "Equine Infectious Anemia Virus (EIAV): What has HIV's country cousin got tell us?", *Veterinary research,* vol. 35, no. 4, pp. 485-512.

Leroux, C., Montelaro, R.C., Sublimec, E. & Cadoré, J.-. (2005) "EIAV (equine infectious anemia virus): To better understand the lentiviral pathogenesis", *Virologie,* vol. 9, no. 4, pp. 289-300.

Leu, Y.-., Rahmatpanah, F., Shi, H., Wei, S.H., Liu, J.C., Yan, P.S. and Huang, T.H.-. (2003) "Double RNA interference of DNMT3b and DNMT1 enhances DNA demethylation and gene reactivation", *Cancer research,* vol. 63, no. 19, pp. 6110-6115.

Levine, A.J. (2009) "The common mechanisms of transformation by the small DNA tumor viruses: The inactivation of tumor suppressor gene products: p53", *Virology,* vol. 384, no. 2, pp. 285-293.

Lewinski, M.K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannenhalli, S., Berry, C.C., Ecker, J.R. &

Bushman, F.D. (2006) "Retroviral DNA integration: Viral and cellular determinants of target-site selection", *PLoS Pathogens,* vol. 2, no. 6, pp. 0611-0622.

Li, C.L., Xiong, D., Stamatoyannopoulos, G. & Emery, D.W. (2009) "Genomic and functional assays demonstrate reduced gammaretroviral vector genotoxicity associated with use of the cHS4 chromatin insulator", *Molecular Therapy,* vol. 17, no. 4, pp. 716-724.

Li, E., Bestor, T.H. & Jaenisch, R. (1992) "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality", *Cell,* vol. 69, no. 6, pp. 915-926.

Li, H., Xu, D., Toh, B.-. & Liu, J.-. (2006) "TGF-β and cancer: Is Smad3 a repressor of hTERT gene?", *Cell research,* vol. 16, no. 2, pp. 169-173.

Li, S., Saragovi, H.U., Racine, R.J. & Fahnestock, M. (2003) "A ligand of the p65/p95 receptor suppresses perforant path kindling, kindling-induced mossy fiber sprouting, and hilar area changes in adult rats", *Neuroscience,* vol. 119, no. 4, pp. 1147-1156.

Lichtenauer-Kaligis, E.G.R., Thijssen, J., Den Dulk, H., Van De Putte, P., Giphart-Gassler, M. & Tasseron-de-Jong, J.G. (1995) "UV-induced mutagenesis in the endogenous hprt gene and in hprt cDNA genes integrated at different positions of the human genome", *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis,* vol. 326, no. 1, pp. 131-146.

Lim, S.-., Rogers, T., Chan, T., Whitney, J.B., Kim, J., Sodroski, J. & Letvin, N.L. (2010) "TRIM5α modulates immunodeficiency virus control in rhesus monkeys", *PLoS Pathogens,* vol. 6, no. 1, art no. e1000738.

Lindahl, T. (1993) "Instability and decay of the primary structure of DNA", *Nature,* vol. 362, no. 6422, pp. 709-715.

Llovet, J.M. & Bruix, J. (2008) "Molecular targeted therapies in hepatocellular carcinoma", *Hepatology,* vol. 48, no. 4, pp. 1312-1327.

Lo, M., Bloom, M.L., Imada, K., Berg, M., Bollenbacher, J.M., Bloom, E.T., Kelsall, B.L. & Leonard, W.J. (1999) "Restoration of lymphoid populations in a murine model of X-linked severe combined immunodeficiency by a gene-therapy approach", *Blood,* vol. 94, no. 9, pp. 3027-3036.

Lohmann, D.R. (1999) "RB1 gene mutations in retinoblastoma", *Human mutation,* vol. 14, no. 4, pp. 283-288.

Lohmann, D.R., Brandt, B., Passarge, E. & Horsthemke, B. (1997) "Molecular analysis and predictive testing in retinoblastoma: Impact on ophthalmologic routine", *Ophthalmologe,* vol. 94, no. 4, pp. 263-267.

Lu, A.L., Clark, S. & Modrich, P. (1983) "Methyl-directed repair of DNA base-pair mismatches in vitro.", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 80, no. 15, pp. 4639-4643.

Lu, D.R., Zhou, J.M., Zheng, B., Qiu, X.F., Xue, J.L., Wang, J.M., Meng, P.L., Han, F.L., Ming, B.H. & Wang, X.P. (1993) "Stage I clinical trial of gene therapy for hemophilia B.", *Science in China.Series B, Chemistry, life sciences & earth sciences,* vol. 36, no. 11, pp. 1342-1351.

MacKenzie, T.C., Kobinger, G.P., Kootstra, N.A., Radu, A., Sena-Esteves, M., Bouchard, S., Wilson, J.M., Verma, I.M. & Flake, A.W. (2002) "Efficient transduction of liver and muscle after in utero injection of lentiviral vectors with different pseudotypes", *Molecular Therapy,* vol. 6, no. 3, pp. 349-358.

Maetzig, T., Galla, M., Baum, C. & Schambach, A. (2011) "Gammaretroviral vectors: Biology, technology and application", *Viruses,* vol. 3, no. 6, pp. 677-713.

Malech, H.L., Maples, P.B., Whiting-Theobald, N., Linton, G.F., Sekhsaria, S., Vowells, S.J., Li, F., Miller, J.A., Decarlo, E., Holland, S.M., Leitman, S.F., Carter, C.S., Butz, R.E., Read, E.J., Fleisher, T.A., Schneiderman, R.D., Van Epps, D.E., Spratt, S.K., Maack, C.A., Rokovich, J.A., Cohen, L.K. & Gallin, J.I. (1997) "Prolonged production of NADPH oxidase-corrected granulocytes after gene therapy of chronic granulomatous disease", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 94, no. 22, pp. 12133-12138.

Manilla, P., Rebello, T., Afable, C., Lu, X., Slepushkin, V., Humeau, L.M., Schonely, K., Ni, Y., Binder, G.K., Levine, B.L., Macgregor, R.-., June, C.H. & Dropulic, B. (2005) "Regulatory considerations for novel gene therapy products: A review of the process leading to the first clinical lentiviral vector", *Human Gene Therapy,* vol. 16, no. 1, pp. 17-25.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F. & Rothberg, J.M. (2005) "Genome sequencing in microfabricated high-density picolitre reactors", *Nature,* vol. 437, no. 7057, pp. 376-380.

Marrero, C.R. & Marrero, J.A. (2007) "Viral Hepatitis and Hepatocellular Carcinoma", *Archives of Medical Research,* vol. 38, no. 6, pp. 612-620.

Martinez, L.A., Goluszko, E., Chen, H.-., Leone, G., Post, S., Lozano, G., Chen, Z. & Chauchereau, A. (2010) "E2F3 is a mediator of DNA damage-induced apoptosis", *Molecular and cellular biology,* vol. 30, no. 2, pp. 524-536.

Martín-Hernández, J., Sørensen, A.B. & Pedersen, F.S. (2006) "Non-identical patterns of proviral insertions around host transcription units in lymphomas induced by different strains of murine leukemia virus", *Virology,* vol. 353, no. 1, pp. 193-199.

Martín-Hernández, J., Sørensen, A.B. & Pedersen, F.S. (2001) "Murine leukemia virus proviral insertions between the N-ras and unr genes in B-cell lymphoma DNA affect the expression of N-ras only", *Journal of virology,* vol. 75, no. 23, pp. 11907-11912.

Matlin, A.J., Clark, F. & Smith, C.W.J. (2005) "Understanding alternative splicing: Towards a cellular code", *Nature Reviews Molecular Cell Biology,* vol. 6, no. 5, pp. 386-398.

Matouk, C.C. & Marsden, P.A. (2008) "Epigenetic regulation of vascular endothelial gene expression", *Circulation research,* vol. 102, no. 8, pp. 873-887.

Mátrai, J., Chuah, M.K.L. & Vandendriessche, T. (2010) "Preclinical and clinical progress in hemophilia gene therapy", *Current opinion in hematology,* vol. 17, no. 5, pp. 387-392.

McCabe, M.T., Davis, J.N. & Day, M.L. (2005) "Regulation of DNA methyltransferase 1 by the pRb/E2F1 pathway", *Cancer research,* vol. 65, no. 9, pp. 3624-3632.

McCabe, M.T., Low, J.A., Daignault, S., Imperiale, M.J., Wojno, K.J. & Day, M.L. (2006) "Inhibition of DNA methyltransferase activity prevents tumorigenesis in a mouse model of prostate cancer", *Cancer research,* vol. 66, no. 1, pp. 385-392.

McGinnis, S. & Madden, T.L. (2004) "BLAST: At the core of a powerful and diverse set of sequence analysis tools", *Nucleic acids research,* vol. 32, no. WEB SERVER ISS., pp. W20-W25.

McKay, T., Patel, M., Pickles, R.J., Johnson, L.G. & Olsen, J.C. (2006) "Influenza M2 envelope protein augments avian influenza hemagglutinin pseudotyping of lentiviral vectors", *Gene therapy,* vol. 13, no. 8, pp. 715-724.

Meissner, F., Seger, R.A., Moshous, D., Fischer, A., Reichenbach, J. & Zychlinsky, A. (2010) "Inflammasome activation in NADPH oxidase defective mononuclear phagocytes from patients with chronic granulomatous disease", *Blood,* vol. 116, no. 9, pp. 1570-1573.

Merlo, A., Herman, J.G., Mao, L., Lee, D.J., Gabrielson, E., Burger, P.C., Baylin, S.B. & Sidransky, D. (1995) "5' CpG island methylation is associated with

transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers", *Nature medicine,* vol. 1, no. 7, pp. 686-692.

Meyer, J., Rhein, M., Schiedlmeier, B., Kustikova, O., Rudolph, C., Kamino, K., Neumann, T., Yang, M., Wahlers, A., Fehse, B., Reuther, G.W., Schlegelberger, B., Ganser, A., Baum, C. & Li, Z. (2007) "Remarkable leukemogenic potency and quality of a constitutively active neurotrophin receptor, ΔTrkA", *Leukemia,* vol. 21, no. 10, pp. 2171-2180.

Miller, A.D. (1996) "Cell-surface receptors for retroviruses and implications for gene transfer", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 93, no. 21, pp. 11407-11413.

Miller, D.G., Adam, M.A. & Miller, A.D. (1990) "Gene transfer by retrovirus vectors occurs only in cells that are actively replicating at the time of infection", *Molecular and cellular biology,* vol. 10, no. 8, pp. 4239-4242.

Miller, J.S., Tessmer-Tuck, J., Blake, N., Lund, J., Scott, A., Blazar, B.R. & Orchard, P.J. (1997) "Endogenous IL-2 production by natural killer cells maintains cytotoxic and proliferative capacity following retroviral-mediated gene transfer", *Experimental hematology,* vol. 25, no. 11, pp. 1140-1148.

Mirsalis, J.C. (1995) "Transgenic models for detection of mutations in tumors and normal tissues of rodents", *Toxicology letters,* vol. 82-83, pp. 131-134.

Mirsalis, J.C., Monforte, J.A. & Winegar, R.A. (1995) "Transgenic animal models for detection of in vivo mutations", *Annual Review of Pharmacology and Toxicology,* vol. 35, pp. 145-164.

Mitchell, R., Chiang, C.-., Berry, C. & Bushman, F. (2003) "Global analysis of cellular transcription following infection with an HIV-based vector", *Molecular Therapy,* vol. 8, no. 4, pp. 674-687.

Mitchell, R.S., Beitzel, B.F., Schroder, A.R.W., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R. & Bushman, F.D. (2004) "Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences", *PLoS Biology,* vol. 2, no. 8, art no. e234.

Miyoshi, H., Blömer, U., Takahashi, M., Gage, F.H. & Verma, I.M. (1998) "Development of a self-inactivating lentivirus vector", *Journal of virology,* vol. 72, no. 10, pp. 8150-5157.

Modlich, U., Bohne, J., Schmidt, M., Von Kalle, C., Knöss, S., Schambach, A. & Baum, C. (2006) "Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity", *Blood,* vol. 108, no. 8, pp. 2545-2553.

Modlich, U., Kustikova, O.S., Schmidt, M., Rudolph, C., Meyer, J., Li, Z., Kamino, K., Von Neuhoff, N., Schlegelberger, B., Kuehlcke, K., Bunting, K.D., Schmidt, S., Deichmann, A., Von Kalle, C., Fehse, B. & Baum, C. (2005) "Leukemias

following retroviral transfer of multidrug resistance 1 (MDR1) are driven by combinatorial insertional mutagenesis", *Blood,* vol. 105, no. 11, pp. 4235-4246.

Mölders, H., Defesche, J., Müller, D., Bonner, T.I., Rapp, U.R. & Müller, R. (1985) "Integration of transfected LTR sequences into the c-raf proto-oncogene: activation by promoter insertion.", *EMBO Journal,* vol. 4, no. 3, pp. 693-698.

Montero, L.M., Filipski, J., Gil, P., Capel, J., Martinez-Zapater, J.M. & Salinas, J. (1992) "The distribution of 5-methylcytosine in the nuclear genome of plants", *Nucleic acids research,* vol. 20, no. 12, pp. 3207-3210.

Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Bartholomae, C.C., Ranzani, M., Benedicenti, F., Sergi, L.S., Ambrosi, A., Ponzoni, M., Doglioni, C., Di Serio, C., Von Kalle, C. & Naldini, L. (2009) "The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy", *Journal of Clinical Investigation,* vol. 119, no. 4, pp. 964-975.

Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Ponzoni, M., Bartholomae, C., Sergi, L.S., Benedicenti, F., Ambrosi, A., Di Serio, C., Doglioni, C., Von Kalle, C. & Naldini, L. (2006) "Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration", *Nature biotechnology,* vol. 24, no. 6, pp. 687-696.

Moolten, F.L. & Cupples, L.A. (1992) "A model for predicting the risk of cancer consequent to retroviral gene therapy", *Human Gene Therapy,* vol. 3, no. 5, pp. 479-486.

Morgan, R.A., Dudley, M.E., Wunderlich, J.R., Hughes, M.S., Yang, J.C., Sherry, R.M., Royal, R.E., Topalían, S.L., Kammula, U.S., Restifo, N.P., Zheng, Z., Nahvi, A., De Vries, C.R., Rogers-Freezer, L.J., Mavroukakis, S.A. & Rosenberg, S.A. (2006) "Cancer regression in patients after transfer of genetically engineered lymphocytes", *Science,* vol. 314, no. 5796, pp. 126-129.

Mortusewicz, O., Schermelleh, L., Walter, J., Cardoso, M.C. & Leonhardt, H. (2005) "Recruitment of DNA methyltransferase I to DNA repair sites", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, no. 25, pp. 8905-8909.

Mothes, W., Boerger, A.L., Narayan, S., Cunningham, J.M. & Young, J.A.T. (2000) "Retroviral entry mediated by receptor priming and low pH triggering of an envelope glycoprotein", *Cell,* vol. 103, no. 4, pp. 679-689.

Müller, H., Bracken, A.P., Vernell, R., Moroni, M.C., Christians, F., Grassilli, E., Prosperini, E., Vigo, E., Oliner, J.D. & Helin, K. (2001) "E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis", *Genes and Development,* vol. 15, no. 3, pp. 267-285.

Müller, H. & Helin, K. (2000) "The E2F transcription factors: Key regulators of cell proliferation", *Biochimica et Biophysica Acta - Reviews on Cancer,* vol. 1470, no. 1, pp. M1-M12.

Muul, L.M., Tuschong, L.M., Soenen, S.L., Jagadeesh, G.J., Ramsey, W.J., Long, Z., Carter, C.S., Garabedian, E.K., Alleyne, M., Brown, M., Bernstein, W., Schurman, S.H., Fleisher, T.A., Leitman, S.F., Dunbar, C.E., Blaese, R.M. & Candotti, F. (2003) "Persistence and expression of the adenosine deaminase gene for 12 years and immune reaction to gene transfer components: Long-term results of the first clinical gene therapy trial", *Blood,* vol. 101, no. 7, pp. 2563-2569.

Nagakubo, D., Taira, T., Kitaura, H., Ikeda, M., Tamai, K., Iguchi-Ariga, S.M.M. and Ariga, H. (1997) "DJ-1, a novel oncogene which transforms mouse NIH3T3 cells in cooperation with ras", *Biochemical and biophysical research communications,* vol. 231, no. 2, pp. 509-513.

Naldini, L., Blomer, U., Gage, F.H., Trono, D. & Verma, I.M. (1996a) "Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 93, no. 21, pp. 11382-11388.

Naldini, L., Blömer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M. & Trono, D. (1996b) "In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector", *Science,* vol. 272, no. 5259, pp. 263-267.

Narayan, O. & Clements, J.E. (1989) "Biology and pathogenesis of lentiviruses", *Journal of General Virology,* vol. 70, no. 7, pp. 1617-1639.

Nevins, J.R. (2001) "The Rb/E2F pathway and cancer", *Human molecular genetics,* vol. 10, no. 7, pp. 699-703.

Nevins, J.R. (1992) "E2F: A link between the Rb tumor suppressor protein and viral oncoproteins", *Science,* vol. 258, no. 5081, pp. 424-429.

Nienhuis, A.W. (2008) "Development of gene therapy for blood disorders", *Blood,* vol. 111, no. 9, pp. 4431-4444.

Nienhuis, A.W., Dunbar, C.E. & Sorrentino, B.P. (2006) "Genotoxicity of Retroviral Integration In Hematopoietic Cells", *Molecular Therapy,* vol. 13, no. 6, pp. 1031-1049.

Nirenberg, M. (2004) "Historical review: Deciphering the genetic code - A personal account", *Trends in biochemical sciences,* vol. 29, no. 1, pp. 46-54.

Nisole, S., Stoye, J.P. & Saïb, A. (2005) "TRIM family proteins: Retroviral restriction and antiviral defence", *Nature Reviews Microbiology,* vol. 3, no. 10, pp. 799-808.

Norton, P.A. (1994) "Alternative pre-mRNA splicing: Factors involved in splice site selection", *Journal of cell science,* vol. 107, no. 1, pp. 1-7.

Nowrouzi, A., Dittrich, M., Klanke, C., Heinkelein, M., Rammling, M., Dandekar, T., von Kalle, C. & Rethwilm, A. (2006) "Genome-wide mapping of foamy virus vector integrations into a human cell line", *Journal of General Virology,* vol. 87, no. 5, pp. 1339-1347.

Oakes, C.C., La Salle, S., Smiraglia, D.J., Robaire, B. & Trasler, J.M. (2007) "Developmental acquisition of genome-wide DNA methylation occurs prior to meiosis in male germ cells", *Developmental biology,* vol. 307, no. 2, pp. 368-379.

Okada, N., Honda, A., Kawabata, M. & Yajima, N. (1997) "Sodium phenobarbital-enhanced mutation frequency in the liver DNA of lacZ transgenic mice treated with diethylnitrosamine", *Mutagenesis,* vol. 12, no. 3, pp. 179-184.

Okano, M., Bell, D.W., Haber, D.A. & Li, E. (1999) "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development", *Cell,* vol. 99, no. 3, pp. 247-257.

Olsen, J.C. (1998) "Gene transfer vectors derived from equine infectious anemia virus", *Gene therapy,* vol. 5, no. 11, pp. 1481-1487.

Olsen, S.K., Brown Jr., R.S. & Siegel, A.B. (2010) "Hepatocellular carcinoma: Review of current treatment with a focus on targeted molecular therapies", *Therapeutic Advances in Gastroenterology,* vol. 3, no. 1, pp. 55-66.

Olson, P., Nelson, S. & Dornburg, R. (1994) "Improved self-inactivating retroviral vectors derived from spleen necrosis virus", *Journal of virology,* vol. 68, no. 11, pp. 7060-7066.

Ory, D.S., Neugeboren, B.A. & Mulligan, R.C. (1996) "A stable human-derived packaging cell line for production of high titer retrovirus/vesicular stomatitis virus G pseudotypes", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 93, no. 21, pp. 11400-11406.

Osten, P., Grinevich, V. & Cetin, A. (2007) "Viral vectors: a wide range of choices and high levels of service.", *Handbook of Experimental Pharmacology,* vol. 178, pp. 177-202.

Ott, M.G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U., Glimm, H., Kühlcke, K., Schilz, A., Kunkel, H., Naundorf, S., Brinkmann, A., Deichmann, A., Fischer, M., Ball, C., Pilz, I., Dunbar, C., Du, Y., Jenkins, N.A., Copeland, N.G., Lüthi, U., Hassan, M., Thrasher, A.J., Hoelzer, D., Von Kalle, C., Seger, R. & Grez, M. (2006) "Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1", *Nature medicine,* vol. 12, no. 4, pp. 401-409.

Overbaugh, J., Miller, A.D. & Eiden, M.V. (2001) "Receptors and entry cofactors for retroviruses include single and multiple transmembrane-spanning proteins as well as newly described glycophosphatidylinositol-anchored and secreted proteins", *Microbiology and Molecular Biology Reviews,* vol. 65, no. 3, pp. 371-389.

Ozturk, M., Arslan-Ergul, A., Bagislar, S., Senturk, S. & Yuzugullu, H. (2009) "Senescence and immortality in hepatocellular carcinoma", *Cancer letters,* vol. 286, no. 1, pp. 103-113.

Pajer, P., Pečenka, V., Králová, J., Karafiát, V., Průková, D., Zemanová, Z., Kodet, R. & Dvořák, M. (2006) "Identification of potential human oncogenes by mapping the common viral integration sites in avian nephroblastoma", *Cancer research,* vol. 66, no. 1, pp. 78-86.

Pannell, D. & Ellis, J. (2001) "Silencing of gene expression: Implications for design of retrovirus vectors", *Reviews in medical virology,* vol. 11, no. 4, pp. 205-217.

Pear, W.S., Nolan, G.P., Scott, M.L. & Baltimore, D. (1993) "Production of high-titer helper-free retroviruses by transient transfection", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 90, no. 18, pp. 8392-8396.

Peebles, D., Gregory, L.G., David, A., Themis, M., Waddington, S.N., Knapton, H.J., Miah, M., Cook, T., Lawrence, L., Nivsarkar, M., Rodeck, C. & Coutelle, C. (2004) "Widespread and efficient marker gene expression in the airway epithelia of fetal sheep after minimally invasive tracheal application of recombinant adenovirus in utero", *Gene therapy,* vol. 11, no. 1, pp. 70-78.

Persons, D.A. (2010) "Editorial: Lentiviral vector gene therapy: Effective and safe?", *Molecular Therapy,* vol. 18, no. 5, pp. 861-862.

Pfeifer, A. & Verma, I.M. (2001) "Gene therapy: Promises and problems", *Annual review of genomic and human genetics,* vol.2, pp.177-211.

Phillips, D.J., Jones, K.L., Scheerlinck, J.-.Y., Hedger, M.P. & De Kretser, D.M. (2001) "Evidence for activin A and follistatin involvement in the systemic inflammatory response", *Molecular and cellular endocrinology,* vol. 180, no. 1-2, pp. 155-162.

Pichavant, C., Aartsma-Rus, A., Clemens, P.R., Davies, K.E., Dickson, G., Takeda, S., Wilton, S.D., Wolff, J.A., Wooddell, C.I., Xiao, X. & Tremblay, J.P. (2011) "Current status of pharmaceutical and genetic therapeutic approaches to treat DMD", *Molecular Therapy,* vol. 19, no. 5, pp. 830-840.

Pierce, A.M., Schneider-Broussard, R., Gimenez-Conti, I.B., Russell, J.L., Conti, C.J. & Johnson, D.G. (1999) "E2F1 has both oncogenic and tumor-suppressive properties in a transgenic model", *Molecular and cellular biology,* vol. 19, no. 9, pp. 6408-6414.

Pike-Overzet, K., De Ridder, D., Weerkamp, F., Baert, M.R.M., Verstegen, M.M., Brugman, M.H., Howe, S.J., Reinders, M.J.T., Thrasher, A.J., Wagemaker, G., Van Dongen, J.J.M. & Staal, F.J.T. (2006) "Gene therapy: Is IL2RG oncogenic in T-cell development?", *Nature,* vol. 443, no. 7109, pp.E6-E7

Poeschla, E.M. & Looney, D.J. (1998a) "CXCR4 is required by a nonprimate lentivirus: Heterologous expression of feline immunodeficiency virus in human, rodent, and feline cells", *Journal of virology,* vol. 72, no. 8, pp. 6858-6866.

Poeschla, E.M., Wong-Staal, F. & Looney, D.J. (1998b) "Efficient transduction of nondividing human cells by feline immunodeficiency virus lentiviral vectors", *Nature medicine,* vol. 4, no. 3, pp. 354-357.

Polager, S. & Ginsberg, D. (2009) "P53 and E2f: Partners in life and death", *Nature Reviews Cancer,* vol. 9, no. 10, pp. 738-748.

Porada, C.D., Tran, N., Eglitis, M., Moen, R.C., Troutman, L., Flake, A.W., Zhao, Y., Anderson, W.F. and Zanjani, E.D. (1998) "In utero gene therapy: Transfer and long-term expression of the bacterial neo(r) gene in sheep after direct injection of retroviral vectors into preimmune fetuses", *Human Gene Therapy,* vol. 9, no. 11, pp. 1571-1585.

Porteus, M.H., Connelly, J.P. & Pruett, S.M. (2006) "A look to future directions in gene therapy research for monogenic diseases.", *PLoS genetics,* vol. 2, no. 9, art no.e133.

Preuss, S. & Pikaard, C.S. (2007) "rRNA gene silencing and nucleolar dominance: Insights into a chromosome-scale epigenetic on/off switch", *Biochimica et Biophysica Acta - Gene Structure and Expression,* vol. 1769, no. 5-6, pp. 383-392.

Qasim, W., Gaspar, H.B. & Thrasher, A.J. (2009) "Progress and prospects: Gene therapy for inherited immunodeficiencies", *Gene therapy,* vol. 16, no. 11, pp. 1285-1291.

Raineri, I. and Senn, H.-. (1992) "HIV-1 promotor insertion revealed by selective detection of chimeric provirus-host gene transcripts", *Nucleic acids research,* vol. 20, no. 23, pp. 6261-6266.

Ralph, S.J., Rodríguez-Enríquez, S., Neuzil, J. & Moreno-Sánchez, R. (2010) "Bioenergetic pathways in tumor mitochondria as targets for cancer therapy and the importance of the ROS-induced apoptotic trigger", *Molecular aspects of medicine,* vol. 31, no. 1, pp. 29-59.

Ramezani, A., Hawley, T.S. & Hawley, R.G. (2008) "Reducing the genotoxic potential of retroviral vectors ", *Methods in molecular biology,* vol.434, pp.183-203.

Ramezani, A., Hawley, T.S. & Hawley, R.G. (2003) "Performance- and safety-enhanced lentiviral vectors containing the human interferon-β scaffold

attachment region and the chicken β-globin insulator", *Blood,* vol. 101, no. 12, pp. 4717-4724.

Recchia, A., Bonini, C., Magnani, Z., Urbinati, F., Sartori, D., Muraro, S., Tagliafico, E., Bondanza, A., Stanghellini, M.T.L., Bernardi, M., Pescarollo, A., Ciceri, F., Bordignon, C. & Mavilio, F. (2006) "Retroviral vector integration deregulates gene expression but has no consequence on the biology and function of transplanted T cells", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 103, no. 5, pp. 1457-1462.

Recio, L. & Goldsworthy, T.L. (1995) "The use of transgenic mice for studying mutagenicity induced by 1,3 butadiene", *Toxicology letters,* vol. 82-83, pp. 607-612.

Reed, C.A., Mayhew, C.N., McClendon, A.K., Yang, X., Witkiewicz, A. & Knudsen, E.S. (2009) "RB has a critical role in mediating the in vivo checkpoint response, mitigating secondary DNA damage and suppressing liver tumorigenesis initiated by aflatoxin B1", *Oncogene,* vol. 28, no. 50, pp. 4434-4443.

Reliene, R. & Schiestl, R.H. (2003) "Mouse models for induced genetic instability at endogenous loci", *Oncogene,* vol. 22, no. 45 REV. ISS. 5, pp. 7000-7010.

Remus, R., Kämmer, C., Heller, H., Schmitz, B., Schell, G. & Doerfler, W. (1999) "Insertion of foreign DNA into an established mammalian genome can alter the methylation of cellular DNA sequences", *Journal of virology,* vol. 73, no. 2, pp. 1010-1022.

Rhee, I., Bachman, K.E., Park, B.H., Jair, K.-., Yen, R.-.C., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., Baylin, S.B. and Vogelstein, B. (2002) "DNMT1 and DNMT3b cooperate to silence genes in human cancer cells", *Nature,* vol. 416, no. 6880, pp. 552-556

RIEKE, W.O. (1962) "The in vivo reutilization of lymphocytic and sarcoma DNA by cells", *The Journal of cell biology,* vol. 13, pp. 205-216.

Robertson, K.D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F.A. & Jones, P.A. (1999) "The human DNA methyltransferases (DNMTs) 1, 3a and 3b: Coordinate mRNA expression in normal tissues and overexpression in tumors", *Nucleic acids research,* vol. 27, no. 11, pp. 2291-2298.

Robertson, K.D. & Wolffe, A.P. (2000) "DNA methylation in health and disease", *Nature Reviews Genetics,* vol. 1, no. 1, pp. 11-19.

Rodenhiser, D. & Mann, M. (2006) "Epigenetics and human disease: Translating basic biology into clinical applications", *CMAJ,* vol. 174, no. 3, pp. 341-348.

Roy, A.L., Sen, R. and Roeder, R.G. (2011) "Enhancer-promoter communication and transcriptional regulation of Igh", *Trends in immunology,* vol. 32, no. 11, pp. 532-539.

Rucker, M., Fraites Jr., T.J., Porvasnik, S.L., Lewis, M.A., Zolotukhin, I., Cloutier, D.A. & Byrne, B.J. (2004) "Rescue of enzyme deficiency in embryonic diaphragm in a mouse model of metabolic myopathy: Pompe disease", *Development,* vol. 131, no. 12, pp. 3007-3019.

Ryu, B.Y., Persons, D.A., Evans-Galea, M.V., Gray, J.T. & Nienhuis, A.W. (2007) "A chromatin insulator blocks interactions between globin regulatory elements and cellular promoters in erythroid cells", *Blood Cells, Molecules, and Diseases,* vol. 39, no. 3, pp. 221-228.

Sadat, M.A., Dirscherl, S., Sastry, L., Dantzer, J., Pech, N., Griffin, S., Hawkins, T., Zhao, Y., Barese, C.N., Cross, S., Orazi, A., An, C., Goebel, W.S., Yoder, M.C., Li, X., Grez, M., Cornetta, K., Mooney, S.D. & Dinauer, M.C. (2009) "Retroviral vector integration in post-transplant hematopoiesis in mice conditioned with either submyeloablative or ablative irradiation", *Gene therapy,* vol. 16, no. 12, pp. 1452-1464.

Sadikovic, B., Al-Romaih, K., Squire, J.A. & Zielenska, M. (2008) "Cause and consequences of genetic and epigenetic alterations in human cancer", *Current Genomics,* vol. 9, no. 6, pp. 394-408.

Saenz, D.T., Teo, W., Olsen, J.C. & Poeschla, E.M. (2005) "Restriction of feline immunodeficiency virus by Ref1, Lv1, and primate TRIM5α proteins", *Journal of virology,* vol. 79, no. 24, pp. 15175-15188.

Saggioro, D., Forino, M. & Chieco-Bianchi, L. (1991) "Transcriptional block of HTLV-I LTR by sequence-specific methylation", *Virology,* vol. 182, no. 1, pp. 68-75.

Saito, K., Kawakami, K., Matsumoto, I., Oda, M., Watanabe, G. & Minamoto, T. (2010) "Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non-small cell lung cancer", *Clinical Cancer Research,* vol. 16, no. 8, pp. 2418-2426.

Santarius, T., Shipley, J., Brewer, D., Stratton, M.R. & Cooper, C.S. (2010) "A census of amplified and overexpressed human cancer genes", *Nature Reviews Cancer,* vol. 10, no. 1, pp. 59-64.

Saulnier, S.O., Steinhoff, D., Dinauer, M.C., Zufferey, R., Trono, D., Seger, R.A. & Hossle, J.P. (2000) "Lentivirus-mediated gene transfer of gp91phox corrects chronic granulomatous disease (CGD) phenotype in human X-CGD cells", *Journal of Gene Medicine,* vol. 2, no. 5, pp. 317-325.

Sauter, S.L. & Gasmi, M. (2000) "FIV vector systems", *Somatic cell and molecular genetics,* vol. 26, no. 1-6, pp. 99-129.

Sauter, S.L., Gasmi, M. & Dubensky Jr., T.W. (2003) "A highly efficient gene delivery system derived from feline immunodeficiency virus (FIV).", *Methods in Molecular Medicine,* vol. 76, pp. 405-432.

Schambach, A., Galla, M., Maetzig, T., Loew, R. & Baum, C. (2007) "Improving transcriptional termination of self-inactivating gamma-retroviral and lentiviral vectors", *Molecular Therapy,* vol. 15, no. 6, pp. 1167-1173.

Schmidt, M., Carbonaro, D.A., Speckmann, C., Wissler, M., Bohnsack, J., Elders, M., Aronow, B.J., Nolta, J.A., Kohn, D.B. & Von Kalle, C. (2003) "Clonality analysis after retroviral-mediated gene transfer to CD34+ cells from the cord blood of ADA-deficient SCID neonates", *Nature medicine,* vol. 9, no. 4, pp. 463-468.

Schmidt, M., Schwarzwaelder, K., Bartholomae, C., Zaoui, K., Ball, C., Pilz, I., Braun, S., Glimm, H. & von Kalle, C. (2007) "High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR)", *Nature Methods,* vol. 4, no. 12, pp. 1051-1057.

Schmidt, M., Schwarzwaelder, K., Bartholomae, C.C., Glimm, H. & von Kalle, C. (2009) "Detection of retroviral integration sites by linear amplification-mediated PCR and tracking of individual integration clones in different samples.", *Methods in molecular biology (Clifton, N.J.),* vol. 506, pp. 363-372.

Schmidt, M., Zickler, P., Hoffmann, G., Haas, S., Wissler, M., Muessig, A., Tisdale, J.F., Kuramoto, K., Andrews, R.G., Wu, T., Kiem, H.-., Dunbar, C.E. & Von Kalle, C. (2002) "Polyclonal long-term repopulating stem cell clones in a primate model", *Blood,* vol. 100, no. 8, pp. 2737-2743.

Schneider, H. & Coutelle, C. (1999) "The case for", *Nature medicine,* vol. 5, no. 3, pp. 256-257.

Schröder, A.R.W., Shinn, P., Chen, H., Berry, C., Ecker, J.R. & Bushman, F. (2002) "HIV-1 integration in the human genome favors active genes and local hotspots", *Cell,* vol. 110, no. 4, pp. 521-529.

Schumacher, A., Kapranov, P., Kaminsky, Z., Flanagan, J., Assadzadeh, A., Yau, P., Virtanen, C., Winegarden, N., Cheng, J., Gingeras, T. & Petronis, A. (2006) "Microarray-based DNA methylation profiling: Technology and applications", *Nucleic acids research,* vol. 34, no. 2, pp. 528-542.

Schumacher, A. & Petronis, A. (2006) "Epigenetics of complex diseases: From general theory to laboratory experiments", *Current topics in microbiology and immunology,* vol 310, pp.81-115.

Seger, R.A. (2008) "Modern management of chronic granulomatous disease", *British journal of haematology,* vol. 140, no. 3, pp. 255-266.

Seger, R.A. (2010) "Chronic granulomatous disease: Recent advances in pathophysiology and treatment", *Netherlands Journal of Medicine,* vol. 68, no. 11, pp. 334-340.

Seggewiss, R., Loré, K., Guenaga, F.J., Pittaluga, S., Mattapallil, J., Chow, C.K., Koup, R.A., Camphausen, K., Nason, M.C., Meier-Schellersheim, M., Donahue,

R.E., Blazar, B.R., Dunbar, C.E. & Douek, D.C. (2007) "Keratinocyte growth factor augments immune reconstitution after autologous hematopoietic progenitor cell transplantation in rhesus macaques", *Blood,* vol. 110, no. 1, pp. 441-449.

Seppen, J., van der Rijt, R., Looije, N., van Til, N.P., Lamers, W.H. & Oude Elferink, R.P.J. (2003) "Long-term correction of bilirubin UDPglucuronyltransferase deficiency in rats by in utero lentiviral gene transfer", *Molecular Therapy,* vol. 8, no. 4, pp. 593-599.

Shafiei, F., Rahnama, F., Pawella, L., Mitchell, M.D., Gluckman, P.D. & Lobie, P.E. (2008) "DNMT3A and DNMT3B mediate autocrine hGH repression of plakoglobin gene transcription and consequent phenotypic conversion of mammary carcinoma cells", *Oncogene,* vol. 27, no. 18, pp. 2602-2612.

Shah, M.Y., Vasanthakumar, A., Barnes, N.Y., Figueroa, M.E., Kamp, A., Hendrick, C., Ostler, K.R., Davis, E.M., Lin, S., Anastasi, J., Le Beau, M.M., Moskowitz, I.P., Melnick, A., Pytel, P. & Godley, L.A. (2010) "DNMT3B7, a truncated DNMT3B isoform expressed in human tumors, disrupts embryonic development and accelerates lymphomagenesis", *Cancer research,* vol. 70, no. 14, pp. 5840-5850.

Shai, E., Palmon, A., Panet, A., Marmary, Y., Sherman, Y., Curran, M.A., Galun, E. & Condiotti, R. (2005) "Prolonged transgene expression in murine salivary glands following non-primate lentiviral vector transduction", *Molecular Therapy,* vol. 12, no. 1, pp. 137-143.

Sharpless, N.E., Alson, S., Chan, S., Silver, D.P., Castrillon, D.H. & DePinho, R.A. (2002) "p16INK4a and p53 deficiency cooperate in tumorigenesis", *Cancer research,* vol. 62, no. 10, pp. 2761-2765.

Sheiness, D. & Bishop, J.M. (1979) "DNA and RNA from uninfected vertebrate cells contain nucleotide sequences related to the putative transforming gene of avian myelocytomatosis virus", *Journal of virology,* vol. 31, no. 2, pp. 514-521.

Shelton, G.H., Grant, C.K., Cotter, S.M., Gardner Hardy Jr., M.B.W.D. & DiGiacomo, R.F. (1990) "Feline immunodeficiency virus and Feline leukemie virus infections and their relationships to lymphoid malignancies in cats: A retrospective study (1968-1988)", *Journal of acquired immune deficiency syndromes,* vol. 3, no. 6, pp. 623-630.

Shi, P.A., Hematti, P., Von Kalle, C. & Dunbar, C.E. (2002) "Genetic marking as an approach to studying in vivo hematopoiesis: Progress in the non-human primate model", *Oncogene,* vol. 21, no. 21 REV. ISS. 2, pp. 3274-3283.

Shirodkar, S., Ewen, M., DeCaprio, J.A., Morgan, J., Livingston, D.M. & Chittenden, T. (1992) "The transcription factor E2F interacts with the retinoblastoma product and a p107-cyclin A complex in a cell cycle-regulated manner", *Cell,* vol. 68, no. 1, pp. 157-166.

Silverman, R.H., Nguyen, C., Weight, C.J. & Klein, E.A. (2010) "The human retrovirus XMRV in prostate cancer and chronic fatigue syndrome", *Nature Reviews Urology,* vol. 7, no. 7, pp. 392-402.

Singh, S., Johnson, J. & Chellappan, S. (2010) "Small molecule regulators of Rb-E2F pathway as modulators of transcription", *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms,* vol. 1799, no. 10-12, pp. 788-794.

Singhal, R., Deng, X., Chenchik, A.A. & Kandel, E.S. (2011) "Long-distance effects of insertional mutagenesis", *PLoS ONE,* vol. 6, no. 1, art no. e15832.

Sinn, P.L., Sauter, S.L. & McCray Jr., P.B. (2005) "Gene therapy progress and prospects: Development of improved lentiviral and retroviral vectors - Design, biosafety, and production", *Gene therapy,* vol. 12, no. 14, pp. 1089-1098.

Sitaram, R.T., Cairney, C.J., Grabowski, P., Keith, W.N., Hallberg, B., Ljungberg, B. & Roos, G. (2009) "The PTEN regulator DJ-1 is associated with hTERT expression in clear cell renal cell carcinoma", *International Journal of Cancer,* vol. 125, no. 4, pp. 783-790.

Sivasubbu, S., Balciunas, D., Amsterdam, A. & Ekker, S.C. (2007) "Insertional mutagenesis strategies in zebrafish", *Genome biology,* vol. 8, no. SUPPL. 1.

Skasko, M. & Kim, B. (2008) "Compensatory role of human immunodeficiency virus central polypurine tract sequence in kinetically disrupted reverse transcription", *Journal of virology,* vol. 82, no. 15, pp. 7716-7720.

Smith, A.J., Bainbridge, J.W. & Ali, R.R. (2009) "Prospects for retinal gene replacement therapy", *Trends in Genetics,* vol. 25, no. 4, pp. 156-165.

Smith, L. & Byers, J.F. (2002) "Gene therapy in the post-Gelsinger era.", *JONA"S healthcare law, ethics and regulation,* vol. 4, no. 4, pp. 104-110.

Soriano, P., Gridley, T. & Jaenisch, R. (1987) "Retroviruses and insertional mutagenesis in mice: proviral integration at the Mov 34 locus leads to early embryonic death.", *Genes & development,* vol. 1, no. 4, pp. 366-375.

Soudais, C., Shiho, T., Sharara, L.I., Guy-Grand, D., Taniguchi, T., Fischer, A. & Di Santo, J.P. (2000) "Stable and functional lymphoid reconstitution of common cytokine receptor γ chain deficient mice by retroviral-mediated gene transfer", *Blood,* vol. 95, no. 10, pp. 3071-3077.

Sourvinos, G., Tsatsanis, C. & Spandidos, D.A. (1999) "Overexpression of the Tpl-2/Cot oncogene in human breast cancer", *Oncogene,* vol. 18, no. 35, pp. 4968-4973.

Spector, D.H., Smith, K. & Padgett, T. (1978) "Uninfected avian cells contain RNA related to the transforming gene of avian sarcoma viruses", *Cell,* vol. 13, no. 2, pp. 371-379.

Spector, D.H., Varmus, H.E. and Bishop, J.M. (1978) "Nucleotide sequences related to the transforming gene of avian sarcoma virus are present in DNA of uninfected vertebrates", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 75, no. 9, pp. 4102-4106.

Stein, C.S., Kang, Y., Sauter, S.L., Townsend, K., Staber, P., Derksen, T.A., Martins, I., Qian, J., Davidson, B.L. & McCray Jr., P.B. (2001) "In vivo treatment of hemophilia A and mucopolysaccharidosis type VII using nonprimate lentiviral vectors", *Molecular Therapy,* vol. 3, no. 6, pp. 850-856.

Stein, L.D. (2004) "End of the beginning", *Nature,* vol. 431, no. 7011, pp. 915-916.

Stein, S., Ott, M.G., Schultze-Strasser, S., Jauch, A., Burwinkel, B., Kinner, A., Schmidt, M., Krämer, A., Schwäble, J., Glimm, H., Koehl, U., Preiss, C., Ball, C., Martin, H., Göhring, G., Schwarzwaelder, K., Hofmann, W.-., Karakaya, K., Tchatchou, S., Yang, R., Reinecke, P., Kühlcke, K., Schlegelberger, B., Thrasher, A.J., Hoelzer, D., Seger, R., Von Kalle, C. & Grez, M. (2010) "Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease", *Nature medicine,* vol. 16, no. 2, pp. 198-204.

Stevens, C. & La Thangue, N.B. (2003) "E2F and cell cycle control: A double-edged sword", *Archives of Biochemistry and Biophysics,* vol. 412, no. 2, pp. 157-169.

Stevens, C. & La Thangue, N.B. (2003) "A new role for E2F-1 in checkpoint control.", *Cell cycle (Georgetown, Tex.),* vol. 2, no. 5, pp. 435-437.

Stiewe, T. & Pützer, B.M. (2001) "p73 in apoptosis", *Apoptosis,* vol. 6, no. 6, pp. 447-452.

Stocking, C., Bergholz, U., Friel, J., Klingler, K., Wagener, T., Starke, C., Kitamura, T., Miyajima, A. & Ostertag, W. (1993) "Distinct classes of factor-independent mutants can be isolated after retroviral mutagenesis of a human myeloid stem cell line", *Growth Factors,* vol. 8, no. 3, pp. 197-209.

Studdert, M.J. (1996) "Equine herpesvirus 2 and disease.", *Equine veterinary journal,* vol. 28, no. 6, pp. 426-428.

Suerth, J.D., Maetzig, T., Galla, M., Baum, C. & Schambach, A. (2010) "Self-inactivating alpharetroviral vectors with a split-packaging design", *Journal of virology,* vol. 84, no. 13, pp. 6626-6635.

Surani, M.A. (1998) "Imprinting and the initiation of gene silencing in the germ line", *Cell,* vol. 93, no. 3, pp. 309-312.

Suzuki, T., Minehata, K.-., Akagi, K., Jenkins, N.A. & Copeland, N.G. (2006) "Tumor suppressor gene identification using retroviral insertional mutagenesis in Blm-deficient mice", *EMBO Journal,* vol. 25, no. 14, pp. 3422-3431.

Takai, D., Yagi, Y., Habib, N., Sugimura, T. & Ushijima, T. (2000) "Hypomethylation of LINE1 retrotransposon in human hepatocellular carcinomas, but not in surrounding liver cirrhosis", *Japanese journal of clinical oncology,* vol. 30, no. 7, pp. 306-309.

Tan, H.H. & Porter, A.G. (2009) "p21WAF1 negatively regulates DNMT1 expression in mammalian cells", *Biochemical and biophysical research communications,* vol. 382, no. 1, pp. 171-176.

Tan, M.H., Smith, A.J., Pawlyk, B., Xu, X., Liu, X., Bainbridge, J.B., Basche, M., McIntosh, J., Tran, H.V., Nathwani, A., Li, T. & Ali, R.R. (2009) "Gene therapy for retinitis pigmentosa and Leber congenital amaurosis caused by defects in AIPL1: Effective rescue of mouse models of partial and complete Aipl1 deficiency using AAV2/2 and AAV2/8 vectors", *Human molecular genetics,* vol. 18, no. 12, pp. 2099-2114.

Tang, H., Kuhen, K.L. & Wong-Staal, F. (1999) "Lentivirus replication and regulation", *Annual review of genetics*, vol. 33, pp. 133-170.

Tannapfel, A. & Wittekind, C. (2002) "Genes involved in hepatocellular carcinoma: Deregulation in cell cycling and apoptosis", *Virchows Archiv,* vol. 440, no. 4, pp. 345-352.

Tao, Q., Huang, H., Geiman, T.M., Lim, C.Y., Fu, L., Qiu, G.-. & Robertson, K.D. (2002) "Defective de novo methylation of viral and cellular DNA sequences in ICF syndrome cells", *Human molecular genetics,* vol. 11, no. 18, pp. 2091-2102.

Tao, Q. & Robertson, K.D. (2003) "Stealth technology: How Epstein-Barr virus utilizes DNA methylation to cloak itself from immune detection", *Clinical Immunology,* vol. 109, no. 1, pp. 53-63.

Tao, Y., Kassatly, R.F., Cress, W.D. & Horowitz, J.M. (1997) "Subunit composition determines E2F DNA-binding site specificity", *Molecular and cellular biology,* vol. 17, no. 12, pp. 6994-7007.

Tarantal, A.F., McDonald, R.J., Jimenez, D.F., Lee, C.I., O'Shea, C.E., Leapley, A.C., Won, R.H., Plopper, C.G., Lutzko, C. & Kohn, D.B. (2005) "Intrapulmonary and intramyocardial gene transfer in rhesus monkeys (Macaca mulatta): Safety and efficiency of HIV-1-derived lentiviral vectors for fetal gene delivery", *Molecular Therapy,* vol. 12, no. 1, pp. 87-98.

Tates, A.D., Grummt, T., Tornqvist, M., Farmer, P.B., Van Dam, F.J., Van Mossel, H., Schoemaker, H.M., Osterman-Golkar, S., Uebel Ch., Tang, Y.S., Zwinderman, A.H., Natarajan, A.T. & Ehrenberg, L. (1991) "Biological and chemical monitoring of occupational exposure to ethylene oxide", *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis,* vol. 250, no. 1-2, pp. 483-497.

Tates, A.D., Van Dam, F.J., Van Mossel, H., Schoemaker, H., Thijssen, J.C.P., Woldring, V.M., Zwinderman, A.H. & Natarajan, A.T. (1991) "Use of the clonal assay for the measurement of frequencies of HPRT mutants in T-lymphocytes from five control populations", *Mutation Research - Environmental Mutagenesis and Related Subjects Including Methodology,* vol. 253, no. 2, pp. 199-213.

Temin, H.M. (1993) "Retrovirus variation and reverse transcription: Abnormal strand transfers result in retrovirus genetic variation", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 90, no. 15, pp. 6900-6903.

Temin, H.M. (1980) "RNA viruses, cancer and development.", *Results and problems in cell differentiation,* vol. 11, pp. 196-201.

Temin, H.M. (1979) "The protovirus hypothesis: Speculations on the significance of RNA-directed DNA synthesis for normal developement and for carcinogenesis", *National Cancer Institute monograph,* vol. No 52, pp. 233-238.

Thacker, E.E., Timares, L. & Matthews, Q.L. (2009) "Strategies to overcome host immunity to adenovirus vectors in vaccine development", *Expert Review of Vaccines,* vol. 8, no. 6, pp. 761-777.

Themis, M., May, D., Coutelle, C. & Newbold, R.F. (2003) "Mutational effects of retrovirus insertion on the genome of V79 cells by an attenuated retrovirus vector: Implications for gene therapy", *Gene therapy,* vol. 10, no. 19, pp. 1703-1711.

Themis, M., Waddington, S.N., Schmidt, M., von Kalle, C., Wang, Y., Al-Allaf, F., Gregory, L.G., Nivsarkar, M., Themis, M., Holder, M.V., Buckley, S.M.K., Dighe, N., Ruthe, A.T., Mistry, A., Bigger, B., Rahim, A., Nguyen, T.H., Trono, D., Thrasher, A.J. & Coutelle, C. (2005) "Oncogenesis following delivery of a nonprimate lentiviral gene therapy vector to fetal and neonatal mice", *Molecular Therapy,* vol. 12, no. 4, pp. 763-771.

Thomas, P. & Smart, T.G. (2005) "HEK293 cell line: A vehicle for the expression of recombinant proteins", *Journal of pharmacological and toxicological methods,* vol. 51, no. 3 SPEC. ISS., pp. 187-200.

Thorgeirsson, S.S., Lee, J.-. & Grisham, J.W. (2006) "Functional genomics of hepatocellular carcinoma", *Hepatology,* vol. 43, no. 2 SUPPL. 1, pp. S145-S150.

Thrasher, A.J., Gaspar, H.B., Baum, C., Modlich, U., Schambach, A., Candotti, F., Otsu, M., Sorrentino, B., Scobie, L., Cameron, E., Blyth, K., Neil, J., Abina, S.H.-., Cavazzana-Calvo, M. & Fischer, A. (2006) "Gene therapy: X-SCID transgene leukaemogenicity", *Nature,* vol. 443, no. 7109, pp. E5-E6.

Tischoff, I. & Tannapfel, A. (2008) "DNA methylation in hepatocellular carcinoma", *World Journal of Gastroenterology,* vol. 14, no. 11, pp. 1741-1748.

Trobridge, G.D., Beard, B.C., Gooch, C., Wohlfahrt, M., Olsen, P., Fletcher, J., Malik, P. & Kiem, H.-. (2008) "Efficient transduction of pigtailed macaque hematopoietic repopulating cells with HIV-based lentiviral vectors", *Blood,* vol. 111, no. 12, pp. 5537-5543.

Trusko, S.P., Hoffman, E.K. & George, D.L. (1989) "Transcriptional activation of cKi-ras proto-oncogene resulting from retroviral promoter insertion", *Nucleic acids research,* vol. 17, no. 22, pp. 9259-9265.

Tsai, C.-., Tsai, C.-., Tse, K.-., Chang, H.-. & Chang, Y.-. (2002) "The Epstein-Barr virus oncogene product, latent membrane protein 1, induces the down-regulation of E-cadherin gene expression via activation of DNA methyltransferases", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, no. 15, pp. 10084-10089.

Tsai, K.Y., Hu, Y., Macleod, K.F., Crowley, D., Yamasaki, L. & Jacks, T. (1998) "Mutation of E2f-1 suppresses apoptosis and inappropriate S phase entry and extends survival of Rb-deficient mouse embryos", *Molecular cell,* vol. 2, no. 3, pp. 293-304.

Tsuchida,K., Nakatani,M., Hitachi,K., Uezumi,A., Sunada,Y., Ageta, H and Inokuchi,K. (2009) "Activin signaling as an emerging target for therapeutic interventions", *Cell Communication and Signaling, vol. 7, no. 15.*

Uren, A.G., Kool, J., Berns, A. & Van Lohuizen, M. (2005) "Retroviral insertional mutagenesis: Past, present and future", *Oncogene,* vol. 24, no. 52, pp. 7656-7672.

Vanin, E.F., Kaloss, M., Broscius, C. & Nienhuis, A.W. (1994) "Characterization of replication-competent retroviruses from nonhuman primates with virus-induced T-cell lymphomas and observations regarding the mechanism of oncogenesis", *Journal of virology,* vol. 68, no. 7, pp. 4241-4250.

Varmus, H.E. (1984) "The molecular genetics of cellular oncogenes.", *Annual Review of Genetics,* vol. 18, pp. 553-612.

Vasseur, S., Afzal, S., Tardivel-Lacombe, J., Park, D.S., Iovanna, J.L. & Mak, T.W. (2009) "DJ-1/PARK7 is an important mediator of hypoxia-induced cellular responses", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 106, no. 4, pp. 1111-1116.

Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E. & Francastel, C. (2010) "Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 107, no. 20, pp. 9281-9286.

Vinciguerra, M. & Foti, M. (2008) "PTEN at the crossroad of metabolic diseases and cancer in the liver", *Annals of Hepatology,* vol. 7, no. 3, pp. 192-199.

Voigt, K., Izsvák, Z. & Ivics, Z. (2008) "Targeted gene insertion for molecular medicine", *Journal of Molecular Medicine,* vol. 86, no. 11, pp. 1205-1219.

von Kalle, C., Fehse, B., Layh-Schmitt, G., Schmidt, M., Kelly, P. & Baum, C. (2004) "Stem cell clonality and genotoxicity in hematopoietic cells: Gene activation side effects should be avoidable", *Seminars in hematology,* vol. 41, no. 4, pp. 303-318.

Von Laer, D., Baum, C. & Protzer, U. (2009) "Antiviral gene therapy", Handbook of experimental pharmacology, vol.189, pp.265-297.

Waddington, C.H. (1942) "The epigenotype", *Endeavour,* vol. 1, pp.18–20.

Waddington, S.N., Nivsarkar, M.S., Mistry, A.R., Buckley, S.M.K., Kemball-Cook, G., Mosley, K.L., Mitrophanous, K., Radcliffe, P., Holder, M.V., Brittan, M., Georgiadis, A., Al-Allaf, F., Bigger, B.W., Gregory, L.G., Terence Cook, H., Ali, R.R., Thrasher, A., Tuddenham, E.G.D., Themis, M. & Coutelle, C. (2004) "Permanent phenotypic correction of hemophilia B in immunocompetent mice by prenatal gene therapy", *Blood,* vol. 104, no. 9, pp. 2714-2721.

Wagner, A.M., Schoeberlein, A. & Surbek, D. (2009) "Fetal gene therapy: Opportunities and risks", *Advanced Drug Delivery Reviews,* vol. 61, no. 10, pp. 813-821.

Walsh, C.E. (1999) "Fetal gene therapy", *Gene therapy,* vol. 6, no. 7, pp. 1200-1201.

Wang, G.P., Ciuffi, A., Leipzig, J., Berry, C.C. & Bushman, F.D. (2007) "HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications", *Genome research,* vol. 17, no. 8, pp. 1186-1194.

Wang, W., Wu, J., Zhang, Z. & Tong, T. (2001) "Characterization of Regulatory Elements on the Promoter Region of p16 INK4a That Contribute to Overexpression of p16 in Senescent Fibroblasts", *Journal of Biological Chemistry,* vol. 276, no. 52, pp. 48655-48661.

Watt, F. & Molloy, P.L. (1988) "Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter.", *Genes & development,* vol. 2, no. 9, pp. 1136-1143.

Weber, E.L. & Cannon, P.M. (2007) "Promoter choice for retroviral vectors: Transcriptional strength versus trans-activation potential", *Human Gene Therapy,* vol. 18, no. 9, pp. 849-860.

Wells, D.J. (2006) "Viral and non-viral methods for gene transfer into skeletal muscle", *Current Opinion in Drug Discovery and Development,* vol. 9, no. 2, pp. 163-168.

Weston, C.R. & Davis, R.J. (2002) "The JNK signal transduction pathway", *Current Opinion in Genetics and Development,* vol. 12, no. 1, pp. 14-21.

Wicker, T., Schlagenhauf, E., Graner, A., Close, T.J., Keller, B. & Stein, N. (2006) "454 sequencing put to the test using the complex genome of barley", *BMC Genomics,* vol. 7.

Wienholz, B.L., Kareta, M.S., Moarefi, A.H., Gordon, C.A., Ginno, P.A. & Chédin, F. (2010) "DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo", *PLoS Genetics,* vol. 6, no. 9.

Wodarz, D. & Levy, D.N. (2007) "Human immunodeficiency virus evolution towards reduced replicative fitness in vivo and the development of AIDS", *Proceedings of the Royal Society B: Biological Sciences,* vol. 274, no. 1624, pp. 2481-2490.

Wolffe, A.P., Jones, P.L. & Wade, P.A. (1999) "DNA demethylation", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 96, no. 11, pp. 5894-5896.

Wong, L.-., Azzouz, M., Walmsley, L.E., Askham, Z., Wilkes, F.J., Mitrophanous, K.A., Kingsman, S.M. & Mazarakis, N.D. (2004) "Transduction patterns of pseudotyped lentiviral vectors in the nervous system", *Molecular Therapy,* vol. 9, no. 1, pp. 101-111.

Wong, W., Billing, J.S., Stranford, S.A., Hyde, K., Fry, J., Morris, P.J. & Wood, K.J. (2003) "Retroviral transfer of donor MHC class I or MHC class II genes into recipient bone marrow cells can induce operational tolerance to alloantigens in vivo", *Human Gene Therapy,* vol. 14, no. 6, pp. 577-590.

Woods, N.-., Bottero, V., Schmidt, M., Von Kalle, C. & Verma, I.M. (2006) "Gene therapy: Therapeutic gene causing lymphoma", *Nature,* vol. 440, no. 7088, pp. 1123.

Wu, X. & Burgess, S.M. (2004) "Integration target site selection for retroviruses and transposable elements", *Cellular and Molecular Life Sciences,* vol. 61, no. 19-20, pp. 2588-2596.

Wu, X., Li, Y., Crise, B. & Burgess, S.M. (2003) "Transcription start regions in the human genome are favored targets for MLV integration", *Science,* vol. 300, no. 5626, pp. 1749-1751.

Wu, Y. (2004) "HIV-1 gene expression: Lessons from provirus and non-integrated DNA", *Retrovirology,* vol. 1.

Wu, Z., Zheng, S. & Yu, Q. (2009) "The E2F family and the role of E2F1 in apoptosis", *International Journal of Biochemistry and Cell Biology,* vol. 41, no. 12, pp. 2389-2397.

Yamashita, H. (2008) "Current research topics in endocrine therapy for breast cancer", *International Journal of Clinical Oncology,* vol. 13, no. 5, pp. 380-383.

Yamashita, K., Kim, M.S., Park, H.L., Tokumaru, Y., Osada, M., Inoue, H., Mori, M. & Sidransky, D. (2008) "HOP/OB1/NECC1 promoter DNA is frequently hypermethylated and involved in tumorigenic ability in esophageal squamous cell carcinoma", *Molecular Cancer Research,* vol. 6, no. 1, pp. 31-41.

Yamashita, S., Takahashi, S., McDonell, N., Watanabe, N., Niwa, T., Hosoya, K., Tsujino, Y., Shirai, T. & Ushijima, T. (2008) "Methylation silencing of transforming growth factor-β receptor type II in rat prostate cancers", *Cancer research,* vol. 68, no. 7, pp. 2112-2121.

Yang, B., Guo, M., Herman, J.G. & Clark, D.P. (2003) "Aberrant promoter methylation profiles of tumor suppressor genes in hepatocellular carcinoma", *American Journal of Pathology,* vol. 163, no. 3, pp. 1101-1107.

Yang, L., Zhu, H.-., Cheng, Z.-., Lu, R., Chen, Y.-. & Fang, J.-. (2005) "Expression and methylation of tumor-associated genes in human gastric cancer cell lines", *World Chinese Journal of Digestology,* vol. 13, no. 13, pp. 1493-1498.

Yee, A.S., Reichel, R., Kovesdi, I. & Nevins, J.R. (1987) "Promoter interaction of the E1A-inducible factor E2F and its potential role in the formation of a multi-component complex.", *EMBO Journal,* vol. 6, no. 7, pp. 2061-2068.

Yu, S.-., Von Ruden, T. and Kantoff, P.W. (1986) "Self-inactivating retroviral vectors designed for transfer of whole genes into mammalian cells", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 83, no. 10, pp. 3194-3198.

Zhao, Z., Wu, Q., Cheng, J., Qiu, X., Zhang, J. & Fan, H. (2010) "Depletion of DNMT3A suppressed cell proliferation and restored PTEN in hepatocellular carcinoma cell", *Journal of Biomedicine and Biotechnology,* vol. 2010.

Zufferey, R., Donello, J.E., Trono, D. & Hope, T.J. (1999) "Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors", *Journal of virology,* vol. 73, no. 4, pp. 2886-2892.

BIBLIOGRAPHY

Clavien.P.A., Breitenstein.S (2009) Malignant liver tumours: Current emerging therapies (3rd Edition) : Jon Wiley & Sons

Coffin.J.M., Huges.S.H.,Varmus.H (1997) Retroviruses, New York: Cold Spring Harbor Laboratory Press

Dorak.M (2006) Real-time PCR BIOS Advanced Methods, London: Taylor & Francis Group

Levinson.W (2006) Review of medical microbiology and immunology : McGraw-Hill Professional

Pallas.V., Sanchez-Nauarro.J., Aparicio.F., James.D (2009) Plant Pathology: Methods in molecular biology: Multiplex Polymerase Chain Reaction (PCR) and Real Time Multiplex PCR For The Simultaneous Deection Of Plant Viruses. Vol 508 pages 193-208: Humana Press

Online resources

Sivalingam,J. and Kon,O.J (2011) " Recent advances and improvements in the biosafety of gene therapy",
http://cdn.intechopen.com/pdfs/17711/InTech-Recent_advances_and_improvements_in_the_biosafety_of_gene_therapy.pdf
accessed online July 2011