An Investigation on the Relationship between DNA Double Strand Breaks in the Mammalian Genome and Gene Therapy Retrovirus Vector Genotoxicity

A Thesis Submitted for the Degree of Doctor of Philosophy

By

Ataun Nasr Bhatti

Department of Life Sciences, Brunel University London

Abstract

Gene therapy uses viral vectors to transfer therapeutic genes to host cells for permanent expression. Retroviral vector (RV) integrase causes double-strand-breaks (DSBs) in the host genome for viral cDNA integration and utilises host repair of these breaks via non-homologous-end-joining (NHEJ). This thesis hypothesises that cells lacking NHEJ are susceptible to retrovirus-mediated genotoxicity.

To determine if presence of the retrovirus genome is required for DSBs, cells were infected with replication defective RV either containing vector genomes or without genomes ("empty"). DSB frequency was measured using immunocytochemistry of phosphorylated histone γH2AX, formed when DSBs occur and eliminated post-NHEJ repair. DSBs were measured in infected cell-lines with NHEJ function (MRC5-SV1) and without (AT5-BIVA and XP14BRneo17) to determine DSB and repair profiles during infection. This showed cells competent for NHEJ have rapid DSB repair whilst cells lacking NHEJ pathways have DSBs remaining high.

Insertional mutagenesis was measured using the hypoxanthine-phosphoribosyl-transferase (HPRT) mutagenesis assay for RV carrying genomes versus empty RV. This analysis relies on *HPRT* knock-out using male V79 and MRC5-SV1 cell lines, carrying one X chromosome, as the *HPRT* gene resides on this chromosome. In each cell-line, empty vectors showed little genotoxic readout compared to the genome carrying particles. However, expression of the viral integrase alone in these cells exhibited the highest mutagenesis.

The potential for genotoxicity by RV due to packaging human-endogenous-retroviruses (HERVs) was evaluated in empty vectors by PCR for HERV-K113 (known to be full length and packaged by retroviruses). Primers designed to amplify a 7.5 kb portion of HERV-K113 gave positive bands, corroborated by restriction digests, and presence in 293T human cells, but not in hamster V79 or mouse 3T3 cells. Further work to determine whether HERV sequence are transferred to infected cells sensitively was not completed in this thesis but discussed as future investigation.

Declaration

I, Ataun Nasr Bhatti, hereby declare that all work presented in this thesis is original and has been prepared by myself, unless otherwise stated. This work has not, and will not, be submitted in whole or in part to another University for the award of any other Degree. The research presented in this thesis was conducted pursuant to the University Code of Research Ethics and all compulsory training required has been completed appropriately.

Acknowledgments

I would like to first and foremost thank God for giving me the strength to complete this PhD.

I would like to thank my principle supervisor, Professor Michael Themis. His guidance and assistance has been crucial for me. I would also like to thank Professor Christopher Parris for providing the cell lines I required and for assisting me with the initial phases of my time point experiments.

I would like to thank Dr. Yaghoub Gozaly Chianea for providing me with assistance and guidance with the ICC and HPRT assay protocols and training. I would also like to thank Dr. Hassan Khonsari.

I would like to thank Dr. Amanda Harvey for providing advice and guidance when I needed it.

I would like to thank Dr. Saqlain Suleman and the rest of my lab group for their assistance and friendship. We have bled together in the trenches of our lab.

I would like to thank my friends Cimankinda Kalala, Felix Wong and Jaskarn Rai for their support.

I would like to thank my mother, my wife and my son for always being there for me and for encouraging me to keep on going.

And finally I would like to thank all my family and relatives who have always remembered me in their prayers.

Dedication

I dedicate this thesis to the two people who would have been the happiest to see me complete my PhD.

My father, Azhar Ahmad Bhatti and my grandfather, Muzaffar Ahmad Bhatti.

Abbreviations

6TG	6-thioguanine
8AG	8-Azaguanine
AAV	Adeno-Associated Virus
ADA-SCID	Adenosine deamninase-deficiency severe combined immunodeficiency
AmpR	Ampicillin Resistance
APOBEC3G	Apolipoprotein B mRNA editing enzyme-catalytic polypeptide-like 3G
BLAST	Basic Local Alignment Search Tool
bp	Base Pairs
CA	Capsid protein
CCD34	Coiled-Coil Domain Containing 34 protein
CCR5	C Chemokine Receptor Type 5
CD4	Cluster of Differentiation 4
CF	Cystic Fibrosis
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
CHM	Choroideremia
CMV	Cytomegalovirus
cPPT	Central Polypurine Tract
CtIP	C-terminal-binding protein interacting protein
CTS	Core Transmembrane Subcomplex
CXCR4	C_X_C Chemokine Receptor type 4
DAPI	4′,6-diamidino-2-phenylindole
dH$_2$O	Distilled Water
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
DNA-PKcs	DNA-dependent kinase catalytic subunit
DSB	Double Strand Break
EDTA	Ethylenediaminetetraacetic acid
EIAV	Equine Infections Anaemia Virus
ERV	Endogenous Retrovirus
F9	Human Factor IX
FBS	Fetal Bovine Serum
FCS	Fetal Calf Serum
FITC	Fluorescein isothiocyanate
FIX	Human Factor IX
FVc	Friend Virus factor c
GFP	Green Fluorescent Protein
γH2AX	Phosphorylated Histone H2AX
GM-CSF	Granulocyte Macrophage Colony Stimulating Factor
H2AX	H2A Histone Family Member X
HAT	Hypoxanthine Aminopterin Thymidine
HCC	Hepatocellular Carcinoma
HERV	Human Endogenous Retroviruses
HIV	Human Immunodeficiency Virus
HPRT	Hypoxanthine-Guanine Phoshoribosyltransferase

HR	Homologous Recombination
HT	Hypoxanthine Thymidine
IL-12	Interleukin-12
IL-2	Interleukin 2
IL2RG	IL2 receptor gamma-chain
IN	Integrase Protein
IR	Infra-Red
IV	Intravenous
kb	Kilo Bases
Lam-PCR	Linear-Amplification Mediated PCR
Lig4	DNA Ligase IV
LMO2	LIM domain only 2
LTR	Long Terminal Repeat
LV	Lentivirus
MA	Matrix Protein
MG^{2+}	Magnesium
MLV	Murine Lukaemia Virus
MOI	Multiplicity of Infection
MoMLV	Moloney Murine Leukaemia Virus
MRE11p	Meiotic Recombination 11 protein
MRN	MRE11p, RAD51p and NBS1 Complex
mRNA	Messenger RNA
NADPH	Nicotine adenine dinucleotide phosphate
NBS1	Nijmegen Breakage Syndrome 1 mutated gene
NC	Nucleocapsid protein
NeoR	Neomycin phosphotransferase
Nef	Negative Factor
NFAT	Nuclear Factor of Activated T-cells
NHEJ	Non-Homologous End Joining
NK	Natural Killer Cells
NLS	Nuclear Localisation Signal
NTRK3	Neurotrophic Receptor Tyrosine Kinase 3
PBS	Phosphate Buffered Saline
P-BS	Primer-Binding Site
PCR	Polymerase Chain Reaction
pH	Potential of Hydrogen
PIC	Pre-Integration complex
PPT	Polypurine Tract
PR	Viral Protease Enzyme
RAD51p	Restriction Site Associated DNA Recombination 51 protein
Rabp	Rat Sarcoma Associated Binding Protein
RCL	Replication Competent Lentivirus
REP-1	Rab escort protein 1
RES	Reticuloendothelial system
RNA	Ribonucleic Acid
RNAseq	RNA Sequencing

RPA	Replication protein A
RRE	Rev responsive element
RSV	Rous Sarcoma Virus
RT	Reverse Transcriptase enzyme
RTC	Reverse Transcription Complex
RTCGD	Retroviral Tagged Cancer Gene Database
RV	Retrovirus
SCID	Severe combined Immunodeficiency
SEM	Standard Error of the Mean
SIN	Self-Inactivating
siRNA	Small interfering RNA
Sp1	Specificity Protein 1
ssDNA	Single Stranded DNA
SU	Surface protein
SV40	Simian virus 40
TAR	Trans-Activation Response
Tat	Trans Activator
TE	Tris-EDTA Buffer
TIL	Tumour Infiltrating Lymphocytes
TM	Transmembrane protein
tRNA	Transfer RNA
T-VEC	Talimogene laherparepvec
U3	Unique in 3' repeat region
U5	Unique in 5' repeat region
UV	Ultra Violet
Vif	Virion infectivity factor
Vpr	Viral Protein R
VSVG	Vesicular stomatitis virus-G
Vpu	Viral Protein U
WHV	Woodchuck Hepatitis Virus
WPRE	Woodchuck Hepatitis Virus post-transcriptional regulatory element
X-ALD	X-linked adrenoleukodystrophy
X-CGD	X-linked Chronic granulomatous disease
XLF	XRCC4-like Factor
X-SCID	X linked SCID
XRCC4	X-ray Repair Cross Complementing 4
Ψ	Encapsidation Signal

Table of Contents

List of Figures

List of Tables

Chapter 1 - Literature Review

1.1 Overview and History of Gene Therapy

Gene therapy is a developing medical technology which aims to treat or prevent a disease caused by a defective or non-functional variant of a gene. This is unlike pharmacological methods which aim to treat the symptoms of diseases without addressing the potential root cause of a condition at the genetic level. Gene therapy functions by inserting the desired version of a gene into cells to overcome the negative effect of a defective existing variant. Gene therapy can also produce a necessary or beneficial protein which is not currently being made by the host genome or is non-functional. The ideal use of gene therapy to treat a condition would allow it to repair or replace the genetic sequence, conserving the location of the correct gene in its natural location and not causing any disruptions to the regulatory material around the altered gene thus preventing any genetic side effects (Crick, 1970; Morange, 2009; Mammen, Ramakrishnan and Sudhakar, 2007; Medline, 2022).

Over several decades, gene therapy has been steadily improving as a viable treatment for various conditions. Gene therapy has been advancing from preclinical studies for various conditions such as haemophilia, cardiovascular disease, HIV and cancer (Fischer and Cavazzana-Calvo, 2007, Crystal and O'Connor, 2006; Levine et al., 2006; Lundstrom, 2023).

Three general types of diseases are targeted for with gene therapy.

- Monogenic conditions which are only affected by one gene being the cause of the disease; for example, cystic fibrosis (CF), which is caused by a mutation in the gene for the CF transmembrane conductance regulator (CFTR) protein. CFTR is utilised in secretions such as sweat, digestive fluids and mucus. The dysfunctional CFTR protein leads to these secretions being much thicker in the lungs, being very difficult to cough out and therefore increasing the frequency of lung infections. The condition is caused by the dysfunction of the *CFTR* gene. The condition is autosomal recessive, meaning both copies of the gene must be defective in order to cause the condition (O'Sullivan and Freedman, 2009). Since a single gene is causing the condition along with the autosomal recessive nature of the condition, theoretically, the introduction of a functional copy of the transgene coding for *CFTR* could improve conditions in a CF patient. However in

practice, the topical application of the *CFTR* transgene in vector constructs directly into the maxillary sinus of patients showed poor efficacy where patients conditions were not significantly improved (Wagner *et al.*, 2002). However, trials have continued try and improve the efficacy of CF treatment via gene therapy; the targeting of around 20-30% of sinus epithelial cells in pig models with CF with adenoviral vectors showed a 7% increase in functional CFTR protein production (potash *et al.*, 2013). Thus monogenic autosomal recessive diseases present a seemingly obvious target for gene therapy, however, the targeting of cells to improve expression may be the greatest hurdle.

- Polygenic conditions which are affected by multiple genes leading to the disease. These diseases are more complex to deal with since the repair of one gene may not be enough to fully treat the condition and thus multiple genes have to be targeted. These conditions are also influenced by external factors such as lifestyle and the environment. This can make it more difficult to treat these conditions as genetics may not be the primary factor affecting the host. Attempts have been made to treat conditions such as cancer and type 1 diabetes using gene therapy. Cancer has been targeted by attempting to introduce transgenes which induce cell death in rapidly growing cells, a strategy similar to chemotherapy (Fung and Gerson, 2004). Type 1 diabetes has been targeted by introducing minicircle DNA preparations with transgenes into hepatocytes that improve gene transcription, mRNA processing and translation. This allowed for increased insulin production upon an increase in glucose concentration (Alam et al., 2013).

- Infectious diseases which are caused by microorganisms directly altering the genome of the host. For example, the human immunodeficiency virus (HIV), which functions by infecting cells and incorporating genetic information into the host's genome. This will allow the virus to hijack the cell machinery when needed for replication (Chan and Kim, 1998). In these cases, gene therapy could be used in multiple ways to prevent infection, for instance, by altering the receptors utilised by the virus, by introducing antisense elements which bind to the viral RNA and disrupts the virus, or by targeting peptides which are involved in fusing during the entry of the virus into the host cell (Conrnu et al., 2021).

Two methods are used for the delivery of genetic material, *in vivo* and *ex vivo*. *In vivo* involves the direct delivery of DNA into the host body. This can be done through injection of DNA into the body or infusion via an IV bag. One of the major *in vivo* methods discussed in this thesis utilising a vector (most commonly viral). This requires for the target cells to be easily accessible by the vector and that the vector will specifically cause expression of the gene in the required cells at the required and safe levels for the correct amount of time without being expressed in non-target cells. Adenoviral vectors are primarily used for these due to their high transfection rates, but the virus has drawbacks of causing an immune response and only allowing short term transgene expression since there is no integration of the material into the host DNA (Mitani and Kubo, 2002; Crystal and O'Connor, 2006; Byrnes et al., 1995).

Ex vivo involves the removal of host cells from the patient, their grown *in vitro* (in culture), their alteration with the therapeutic gene being inserted into them and finally the reinsertion of the cells back to the host's target tissue. This allows for long term expression since the cells act as engineered secretory tissue which can constantly produce the correct protein to the local environment of the cells (Selkirk, 2004).

In the 1960's, Borenfreund and Bendich showed the incorporation of foreign DNA into the nuclei of mammalian cells, 6-24 hours following the introduction of the DNA (Borenfreund and Bendich, 1961). This experiment, along with some others in the 1960's, showed the general characterisation and delivery of certain therapeutic genes, leading to an increased interest in gene therapy (Rieke, 1962; Borenfreund and Bendich, 1961).

The first viral system for gene delivery was developed by the Verma group in 1983. They were able to use a retrovirus to transfer the human gene for *hypoxanthine phosphoribosyltransferase* (*HPRT*) into mice with low *HPRT* (Miller et al., 1983). In 1985, the first in vivo transgene expression in an animal model was done by transferring the *neomycin resistance* gene (*NeoR*) into haematopoietic progenitor cells from the bone marrow of mice which were immune competent. Retroviral expression vector N2 was used. These cells were then grafted into lethally irradiated mice and the vector DNA was detected 10 days following injection (Eglitis et al., 1985).

Gene transfer was done in humans first in 1989 in order to treat patients with advanced metastatic melanoma. Human tumour infiltrating lymphocytes (TIL) were modified utilising retrovirus vectors. The cells were then infused into the patients and were

tracked with the *neomycin* gene. This study was able to demonstrate the transduction of a gene using a retroviral vector (Rosenberg et al., 1990).

The first gene therapy trial was approved and conducted in 1990 where two children with adenosine deamninase-deficiency severe combined immunodeficiency (ADA-SCID) were treated. The normal *adenosine deaminase* gene was introduced into the patients' own T cells *ex vivo* using a retroviral vector. They attempted it on autologous cells to prevent acute graft versus host disease caused by immune transplants (Aiuti et al., 2002). The ADA-SCID lymphocyte trial displayed no adverse effects and a significant expression of ADA was observed in cells recovered from the patients. However, the treatment was not considered optimal because there was only transient expression of the transgene. The trial was still considered useful since it displayed that gene therapy could be utilised for some patients with the disease (Blaese et al., 1995). Despite the increasing number of clinical gene therapy trials in the 1990's, a case showing ideal and lasting clinical benefit was not seen (Scollay, 2001). This was attributed to a number of obstacles including inadequate gene delivery systems, immune response to the vectors and the modified cells, and poor transgene expression (Nathwani et al., 2005).

The first gene therapy trial which led to severe adverse effects used a protocol which utilised an adenoviral vector to treat ornithine transcarbamylase deficiency, a metabolic liver disorder. The 18 year old patient Jesse Gelsinger reacted strongly to the adenovirus-5-serotype vector which caused a severe systemic inflammatory response which led to multiple organ failure and death only days after the vector was administered (Raper et al., 2003).

The French trial for X-linked severe combined immunodeficiency (X-SCID) conducted in 2000 is considered the first clear successful gene therapy trial (Gaspar et al., 2004). The trial involved the removal of bone marrow from ten infants followed by selecting the cells for the CD34 marker to ensure a high amount of haematopoietic progenitors. They were then modified ex vivo with the retroviral vector with the interleukin-2 receptor common gamma chain followed by reintroduction into the patients intravenously. The results showed a recovery to normal T-lymphocyte count within months of treatment along with antigen specific response to immunisation in all patients except one. The therapy was successful at repairing the immune system to the point where immunoglobulin therapy was not required in a majority of the cured patients. A similar trial in Britain showed similar immune system recovery in ten patients (Hacein-Bey-Abina et al., 2003). However, the French trial caused T-cell leukaemia-like expansion in four patients, as

well as one in the British trial. This led to the death of one patient in the French trial (Hacein-Bey-Abina et al., 2003, Howe et al., 2008, Hacein-Bey-Abina et al., 2008). It is thought that the insertion of the retrovirus close to proto-onco genes, such as LMO2, caused the enhancer within the retroviral vector to dysregulate the expression of these proto onco-genes (Aiuti et al., 2007).

In 2006, T cell receptors genes were successfully incorporated ex vivo in autologous lymphocytes via retroviral vectors in metastatic melanoma patients. This improved the engraftment of peripheral blood necessary to improve host lymphocyte levels following immune-depletion (Morgan et al., 2006).

Adeno associated virus (AAV) vectors were used to infuse the *human factor IX* (*FIX*) gene in patients with haemophilia B via peripheral vein injection. This led to increase *FIX* expression allowing for improved bleeding response with minimal side effects (Nathwani et al., 2011). AAV vectors have also been used to proffer the *CHM* gene in patients with choroideremia to partially revert blindness. This gene codes for the Rab escort protein 1 (REP-1), without which the retina is unable to function and slowly degrades and dies off, leading to blindness (Scholl and Sahel, 2014). A herpes simplex virus, Talimogene laherparepvec (T-VEC), was modified to act as a vector to replicate selectively in tumours to produce granulocyte macrophage colony stimulating factor (GM-CSF) to improve antitumor response in melanoma patients (Andtbacka et al., 2015).

In 2018, gene therapy was used to reverse hypertriglyceridemia in mice which have had their lipoprotein lipase knocked out in the striated muscle. Human lipoprotein lipase was transferred into these mice using a retroviral vector. This significantly reduced the amounts of triglycerides in the blood compared to that of untreated mice (Gadek et al., 2018).

As of 2018, 2579 gene therapy trials have been undertaken in 38 countries from the year 1989. This has been a sharp increase as seen from there being 1843 trials in 31 countries in 2012. While a majority of these are still in the earlier phases (I and I/II), the number of trials entering the late phases has also been improving. Early phase trials have provided a better proof of concept for the technique, but most trials have (Ginn et al., 2013) led to an unsatisfactory quantity of cells being modified for the therapeutic effect. However, there have also been a few exceptions which have had more desirable results, for instance the treatment of X-linked adrenoleukodystrophy (X-ALD) (Cartier et al., 2009). The results so far show that gene therapy is at a definite incline in

advancement (Ginn et al., 2018), and as such it is important to ensure that the procedure is safe for the host.

1.2 Vectors

A hurdle to gene therapy is the existence of natural barriers in place to prevent DNA access to the cells.
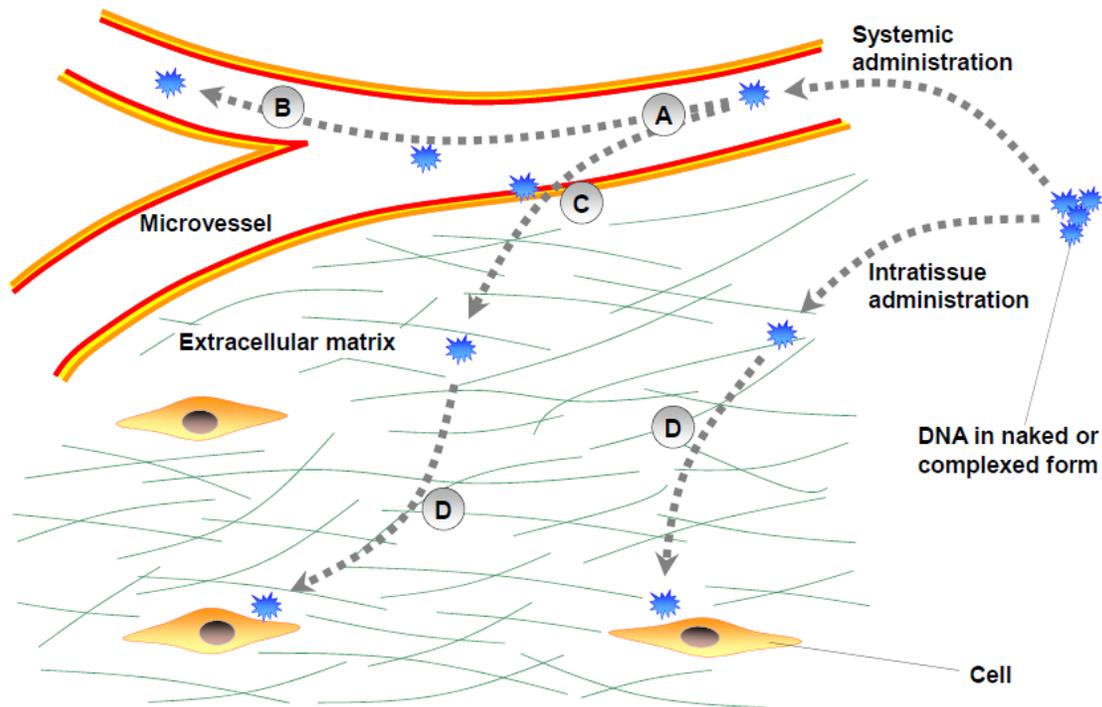
Figure 1.1 Extracellular Barriers to DNA Entry. In order to enter the nucleus and to be expressed for the purposes of gene therapy, DNA must cross extracellular barriers first. These include **A**) Plasma Degradation: DNAses present in the plasma can simply break down DNA. **B**) Reticuloendothelial system (RES): The RES system can clear out DNA from the blood as a part of the immune system. **C**) Microvessel Wall: The physical barrier separating blood from tissue needs to be bypassed for DNA to approach the cells. **D**) Extracellular Matrix: Once through the microvessel wall, the DNA must still get past the extracellular matrix. This is the only barrier which cannot be bypassed via direct intratissue administration (Wang *et al.*, 2013).

Furthermore, these cells themselves have barriers to prevent the entry of foreign genetic material. This becomes particularly troublesome when high gene transfer into the host cells is required.
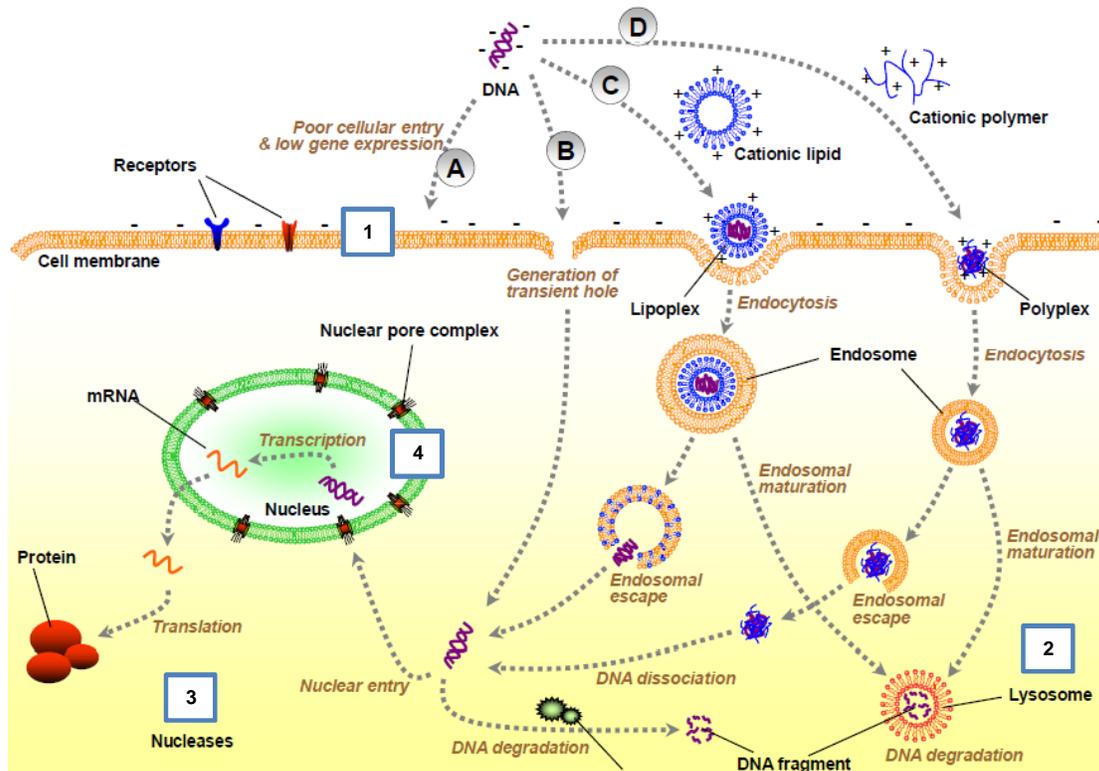
Figure 1.2 Intracellular Barriers to DNA Entry. Illustration of some of the non-viral entry methods used to attempt DNA entry into the nucleus and the barriers in place to prevent this. **A) Direct entry** to the cell is prevented by **1) the cell membrane** itself, a physical barrier. **B)** The most direct physical method to allow this entry is to **generate transient holes** in the cell membrane. **C) Cationic lipids** and **D) cationic polymers** both complex with the cell membrane to form an endosome. This is the most reliable method of entry, however, it leaves the DNA vulnerable to endosomal maturation into **2) lysosomes** which will fill with lysosymes and break down the DNA. If the DNA manages to leave the endosome, or if the naked DNA itself enters through a transient hole in the cell membrane, DNA degradation can still occur due to the presence of **3) nucleases** in the cell. If the DNA manages to get past all these barriers, it must still traverse the **4) nuclear envelope** via the nuclear pore complex in order to be expressed. DNA may also incorporate with the chromatin while the cell is dividing during mitosis (Adapted from Wang *et al.*, 2013).

The virus has been a crucial factor for getting past the barriers for the purposes of gene therapy. This is due to their nature as naturally occurring parasites which bypass the natural barriers of host cells to deliver their genetic material. This ability has been crucial as a tool for gene therapy; so much so, that vectors are classed into two subsets, non-viral and viral.

1.2.1 Non-Viral Vectors

Non-viral vectors include naked DNA, particle based vectors and chemical based vectors (Ramamoorth and Narvekar, 2015). Non-viral vectors lack the natural protection a virus provides to the genetic material they carry. This makes them potentially more vulnerable to extracellular barriers. Physical methods attempt to directly bypass the barriers in place while chemical methods attempt to interact with the cell membrane to form endosomes to allow DNA entry into the cell.

Examples of physical methods are microinjection, injections directly into tissue, jet injection and electroporation. Microinjection is a technique that allows direct penetration of the cell membrane and/or nuclear envelope to introduce the DNA into a single cell at the microscopic level. This is done with a micropipette with a needle 0.5-5 μm in diameter and a pipette to hold the cell in place via suction.
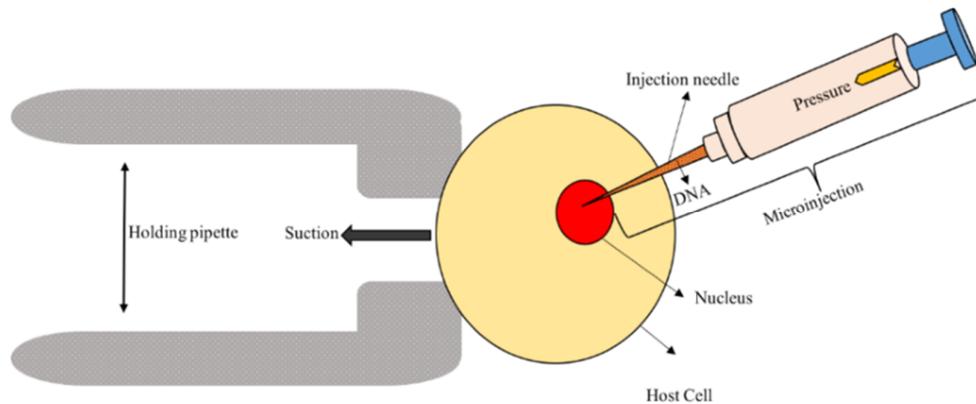
Figure 1.3 Microinjection. The illustration shows an individual cell being held in place by a pipette via suction following which a microinjection with a narrow needle directly pierces the host cell nucleus. The injection then deposits the therapeutic DNA into the cell. This cell can then be reintroduced into the host body. (Adapted from Sari and Iskender, 2022).

It is an efficient method and allows for a large amount of DNA to be transferred. However, it requires individual cell manipulation of cells making it poor in terms of cost efficiency. Transgene expression and persistence are also low (Barber, 1911, Wolf *et al.*, 1990, Davis *et al.*, 1993; Wang *et al.*, 2013).

Injections function similarly to microinjections, however the therapeutic DNA is injected into tissue rather than individual cells. This method has been used to administer naked DNA coding for the *Interleukin-12* (*IL-12*) cytokine and demonstrated transgene expression. It has also been used to introduce naked DNA into muscles while showing transgene expression. The injection method is thus more cost effective as it does not require the formation of vectors. This method can also be used to introduce other

genetic therapeutic agents such as siRNA and DNA/polycation complexes. However, as mentioned, this method has the issue of low transgene expression primarily due to the physical barriers in place listed in **Figure 1.1** (Lui *et al.*, 2001; Herweijer and Wolff, 2003; Thompson and Patel, 2009; Wang *et al.*, 2013). Jet injection is a method that uses a pressurised gas in place of a needle to target cells. This prevents some of the DNA being left in the needle (Ren *et al.*, 2002). It also allows the pressure to be adjusted depending on the tissue being targeted. This method has shown positive results for cancer therapy (Stein *et al.*, 2008; Walther *et al.*, 2008). There have been minor side effects seen with their usage including hyperaemia and superficial side effects such as minor bleeding (Lysakowski *et al*, 2003; Wang *et al.*, 2013).

Electroporation uses electricity to induce the formation of nano-pores in the cell membrane allowing negatively charged DNA to enter following injection of DNA into the tissue surrounding the cells (Neumann *et al.*, 1982).
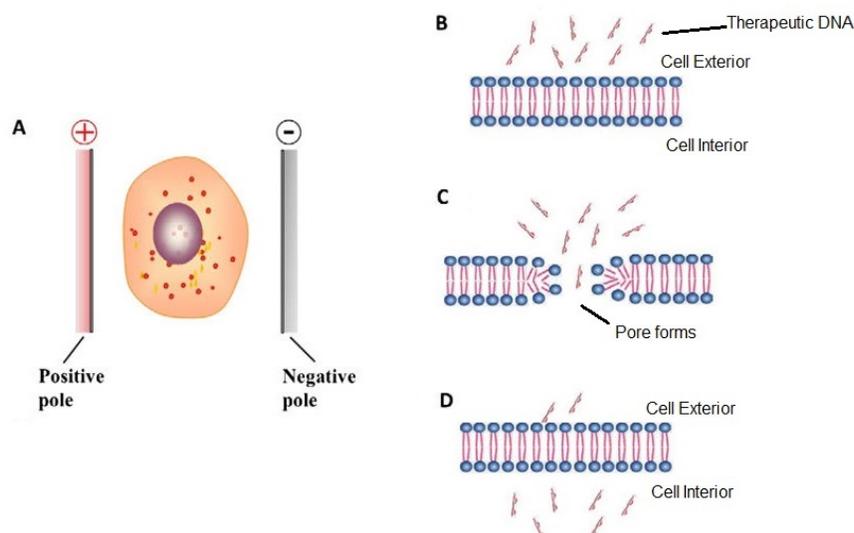
Figure 1.4 Electroportation. The illustration shows how electroportation is used to add therapeutic DNA into host cells. **A**) The cells are place between two electrodes and a pulse of electricity is applied to them. **B**) Before the pulse is applied, the therapeutic DNA is unable to cross the cell membrane. **C**) Once the pulse is applied, it forms pores in the cell membrane, through which the therapeutic DNA can enter. **D**) After some time, the cell is able to heal the damage to the cell membrane and reseal the pore, at which point therapeutic DNA has already entered the cell. (Adapted from Du *et al.*, 2018).

Electroportation has been used to deliver the plasmid IL-12 into tumours of patients with metastatic melanoma (Daud *et al.*, 2008). The results were safe, reproducible and efficient (Wang *et al.*, 2013). In this way, the method is useful in treating tumours. Electroporation does have drawbacks in cases where cell death can occur due to irreversible electroporation. This prevents the pores formed in the cell membrane from

healing and closing up, allowing material to freely move in and out of the cell preventing cell homeostasis (Bolhassani, Khavari and Orafa, 2014, Rubinsky, 2007).

Examples of chemical gene delivery methods are cationic lipids and cationic polymers. Cationic lipids are lipid capsules with hydrophilic heads, hydrophobic tails and linkers to combine them. They house DNA at their core when used as a DNA delivery vector. The positive charges on the surface of the lipid interacts with the negatively charged cell membrane. This interaction causes the lipid to get endocytosed into the cell within an endosome. The DNA must then escape from the endosome. This is achieved due to the cationic nature of the cationic lipids, which interact with the anionic lipids of the endosome and begin to "mix". This mixing will allow the DNA to escape into the cytoplasm (Chesnoy and Huang, 2000; Parker *et al.*, 2003; Hashida *et al.*, 2005; Medina-Kauwe, Xie and Hamm-Alvarez, 2005).
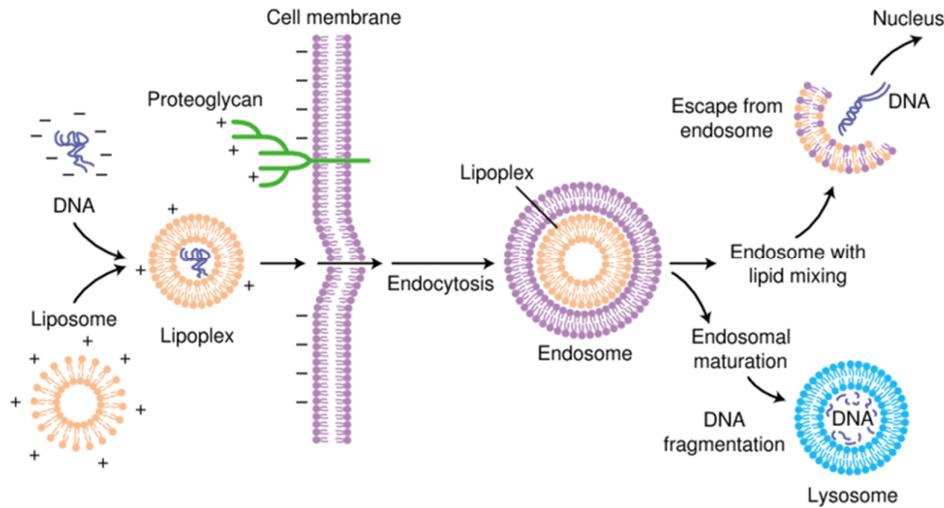
Figure 1.5 Cationic Lipid Principle. The illustration shows how cationic lipids are used to introduce therapeutic DNA into cells. The lipids form a spherical shape known as a liposome. The hydrophilic heads of the lipids face outside and the hydrophobic tails face towards each other to form a bilayer. Negatively charged DNA can interact with the liposome and be engulfed forming a lipoplex. The lipoplex can interact with proteoglycan on the host cell membrane which will the lipoplex to be endocytosed into the cell. The lipoplex is contained in an endosome. The negatively charged lipoplex can interact with positively charged endosome to cause mixing (**figure 1.6**), freeing the DNA inside which can then move towards the nucleus. There is a chance the lipoplex is unable to mix with the endosome, causing the DNA within to be degraded (Parker *et al.*, 2003). (Adapted from Parker *et al.*, 2003).
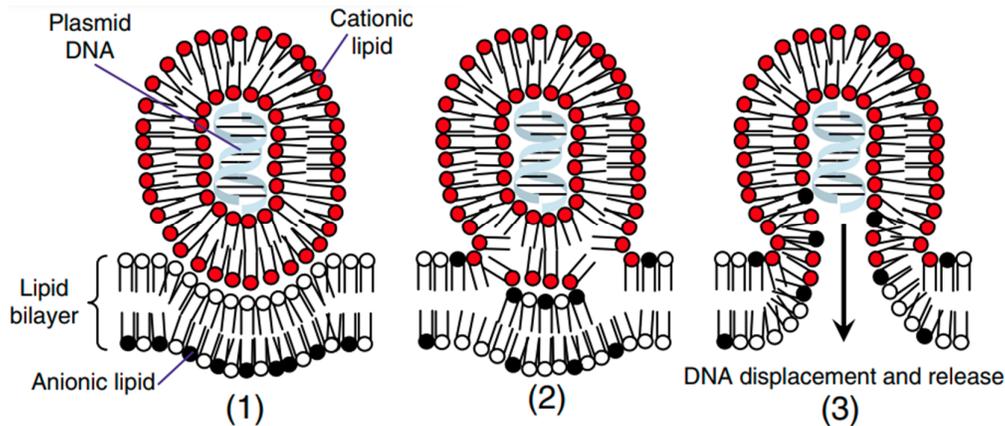
Figure 1.6 Cationic Lipid Mixing. The illustration shows how cationic lipolexes (**figure 1.5**) containing DNA can interact with the endosome to cause release of the DNA. **1**) The cationic lipids of the lipoplex have a positive charge and the anionic lipids of the endosomes have a negative charge. **2**) The cationic lipoplex lipids interact with the endosomal anionic lipids which causes the wall to destabilize. **3**) As the cationic and anionic lipids begin to "mix" into each other, the endosomal wall breaks down allowing the DNA within to escape (Parker *et al.*, 2003; Medina-Kauwe, Xie and Hamm-Alvarez, 2005). (Adapted from Medina-Kauwe, Xie and Hamm-Alvarez, 2005).

Cationic lipids are useful since they are inexpensive, easy to prepare and they can be modified to better target specific cells. However, they have poor transfection efficiency and can be mildly toxic (Parker et al., 2003; Wang *et al.*, 2013).

Cationic polymers are made using amine groups packed at a high density and can be charged with protons when at a neutral pH. The positive charge lets them complex with negatively charged DNA through electrostatic interaction. They are generally more stable and larger than cationic lipids and function with a similar principle by interaction with the cell membrane and being endocytosed. The mechanism through which the DNA escapes the endosome is different. Cationic polymers such as polyethylenemine (PEI), due to their high density of amino groups, can buffer the acidic pH of the endosome interior. This causes osmotic swelling of the endosome leading to rupture, allowing the DNA to escape into the cytoplasm. This is also known as the proton sponge effect (Jin *et al.*, 2014).

The primary advantage of non-viral delivery is the ability to be able to delivery very large DNA molecules into mammalian cells. However, they have also exhibited somewhat poor transfection efficiency and toxicity (Wang *et al.*, 2013). The use of both cationic polymers and liposomes together has shown improved transfection efficiency. The process involves the formation of the cationic polymer which is then complexed with a

liposome. This provides added protection for the cationic polymer within the liposome (Tan *et al.*, 2002; Ueno *et al.*, 2002; Wang *et al.*, 2013).

1.2.2 Viral Vectors

Viruses are a very tempting alternative to non-viral vectors due to their innate abilities to bypass the natural barriers of host cells to deliver their genetic material. However, they potentially lack the safety aspect concordant with non-viral vectors. Even though both viral and non-viral vectors have been used for gene delivery in clinical trials, viral vectors account for around 70% of these. Adenoviral and retroviral vectors have been used primarily (Kay *et al.*, 2001, Glover *et al.*, 2005; Zhao *et al.*, 2008). Viral vectors are chosen based on the cell type that is being altered as well as the duration the expression of the transgene is needed for (Gillet *et al.*, 2009). Viral vectors require certain characteristics to be effective; these include the ability to be propagated reliably and reproducibly, to be purified to high titre and to be able to deliver the gene to the host cell without causing detrimental toxicity (Wanisch and Yáñez–Muñoz, 2009; Yáñez – Muñoz *et al.*, 2006).

1.2.2.1 Retroviruses

Retroviruses are enveloped RNA viruses of the retroviridae family. They are characterised by the general life cycle which involves the reverse transcription of their single stranded RNA genome into double stranded DNA (called the pro virus) which is then integrated into the host cell genome. Each virus particle (virion) is around 100nm in size with a small RNA genome of around 7-12 kb. The *retroviridae* family is categorised based on differences in genetic structure, size of assembled virion, morphology and genome complexity and are divided into seven genera. Five of these have simple structures and oncogenic potential: *Alpharetrovirus, Betaretrovirus, Gammaretrovirus, Deltaretrovirus* and *Epsilonretrovirus*. Two of them are more complex: *Lentivirus* and *Spumavirus*. For the purposes of gene therapy, *Gammaretroviruses*, *Lentiviruses* and *Spumaviruses* are mainly utilised (Baum *et al.*, 2006). Retroviruses have three major aspects which make them useful for gene therapy:

- Uptake of genetic information directly into host cells via receptor mediated mechanisms

- Conversion of single stranded RNA into double stranded DNA which can be integrated into host chromatin

- The incorporation of complete retroviral mRNA being incorporated as a form of genetic information

Retrovirus Biology

While most viruses involve the transcription of DNA into RNA followed by translation into proteins, retroviruses function differently where RNA is first reverse transcribed into DNA, integrated into a host genome, then utilises the host cell machinery to transcribe and translate its genetic material. The retrovirus functions in two general forms, the virion and the provirus. Virions are 100 nm diameter enveloped particles which contain single-stranded RNA. Once the RNA is reverse transcribed and integrated into the host genome, it is called the provirus and contains the coding and non-coding regulatory *cis* acting gene regulatory sequences. The provirus will allow for the expression of the viral proteins followed by the packaging of the material into a virion via budding at the host cell membrane (Watts *et al.*, 2009; Weiss, 2006; Ruscetti, 1995).

Coding Sequences

The retrovirus genome consists of four genes *gag, pro, pol,* and *env.* While simple retroviruses only contain these genes, complex viruses contain more regulatory and accessory genes (Eisenman and Vogt, 1978). The retroviral genome is expressed as a polyprotein made from all four *gag, pro, pol,* and *env* genes. During the assembly of the virion, the polyprotein is cleaved to form the individual proteins required by the virus. The *gag* gene encodes the primary structural proteins of the virion including the matrix (MA), capsid (CA) and nucleocapsid (NC) proteins. The *pol* gene forms the enzymes of the virus including reverse transcriptase (RT) and integrase (IN) which reverse transcribe the viral RNA into double-stranded DNA and the integration of this DNA into the host genome, respectively. The *env* gene encodes the surface (SU) protein and transmembrane (TM) which form the viral envelope proteins involved in forming the virion lipid bilayer. The *pro* gene encodes for the viral protease (PR) which processes the viral polypeptides formed by the other genes of the retrovirus *(Kay et al., 2001).*

Non-Coding Sequences

The flanks of the provirus sequence contain the long terminal repeats (LTR) regions which are divided into the unique in 3' (U3) repeat and the unique in 5' (U5) sequence (Tipper, Cingöz and Coffin, 2012). The U3 region is around 450 nucleotides and acts as a promoter for viral transcripts. The R region is around 100 nucleotides long and has a polyadenylation signal which allows for the processing of viral RNA. The two R regions require homology to allow transcription of the genome. The U5 region is a guanine and u uracil (GU) rich region of about 80 nucleotides which improves the polyadenylation signal in the R region (Huthoff *et al.*, 2003). There are other *cis* acting regulatory sequence sites such as the primer-binding site (P-BS), complementary to a region on the tRNA derived from the host to which it will bind to act as a primer for the minus strand DNA synthesis in reverse transcription (Verma *et al.*, 1971, Dahlberg *et al.*, 1974). The encapsidation signal (Ψ) is found downstream of the PBS and allows the encapsidation of virions. This region is removed during splicing and thus ensures only unspliced, full length transcripts are packaged into virions (Mann and Baltimore, 1985). In HIV-1 (a lentivirus), the polypurine tract (PPT) at 16 nucleotides long is upstream of the U3 and functions as a primer for the plus strand DNA reverse transcription (Smith, Cywinski and Taylor, 1984; Finston and Champoux, 1984). A second plus strand DNA reverse transcription region can also be found in the integrase region within lentiviruses called the central polypurine tract (cPPT) (Charneau *et al.*, 1994).

Virus Structure

The envelope of the retroviral virion is obtained from the host cell membrane during encapsidation. The envelope is bound with the matrix (MA) protein. The transmembrane (TM) region spans the lipid bilayer and anchors the surface receptor (SU) which protrudes outward from the virion as part of the glycoprotein (Perez *et al.*, 1987). The matrix houses the dense virion core which contains the nucleocapsid (NC) capsid complex. The virion also contains the reverse transcriptase (RT) and integrase (IN) proteins. The capsid (CA) proteins are found on the surface of the core and number from around 2000-4000 copies (Halwani *et al.*, 2004).
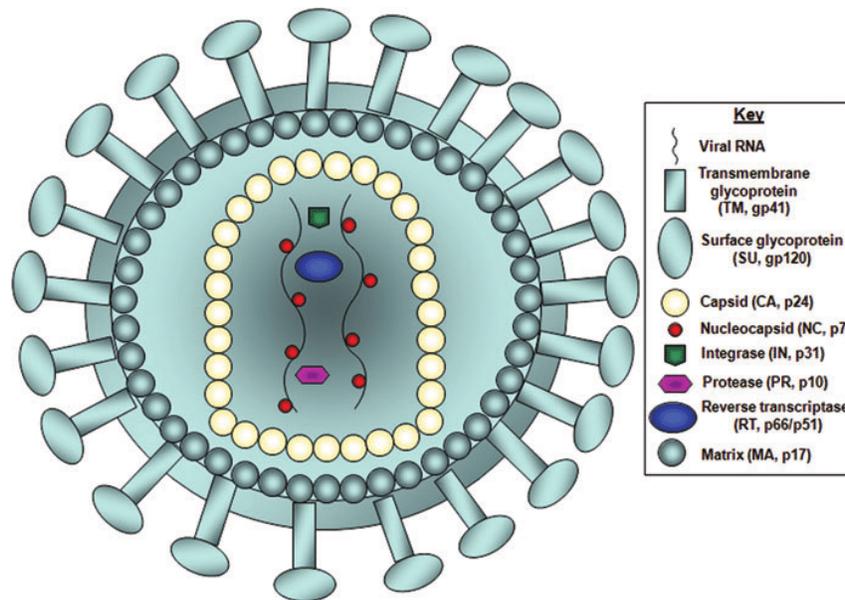
Figure 1.7 General Structure of the HIV-1 Virion. Viral proteins are indicated in the diagram. The Transmembrane glycoprotein (TM) tethers the surface glycoprotein (SU) to the lipid bilayer matrix (MA). The SU is involved in virus signaling with host cells to allow virus entry into the host cell. The matrix (MA) protects the viral genome as well as facilitating viral entry into the host cell. Proteases (PR) are enzymes which are involved in cleavage of viral polyprotein precursors and host cell proteins. The capsid (CA) proteins form a protective barrier around the viral genome. The nucleocapsid (NC) will attach to the viral RNA and will cause condensing for efficient chaperone activity in the host cell. The single stranded RNA genome (RNA) will code for the viral proteins required for infection and replication and will be reverse transcribed for integration into the host genome. The integrase (IN) is an enzyme which will form DSBs in host DNA to allow for integration. Reverse transcriptase (RT) is a DNA polymerase enzyme which will convert viral RNA to double stranded cDNA which will then be integrated into the host genome. Image modified from Hutson *et al*., 2014 (Baum et al., 2006; Halwani *et al*., 2004).

1.2.2.2 Lentiviruses

Lentiviruses are able to deliver large and complex transgenes of up to 10 kb to target cells and as such are ideal for gene therapy (Verma and Weitzman, 2005). In addition to the *gag, pol* and *env* genes, lentiviruses carry regulatory genes, *tat* and *rev*, and auxiliary genes, *vpr, nef, vpu* and *vif*, and are involved in regulating the synthesis and processing of the viral DNA while allowing viral replication (Das and Jameel, 2005; Pfeifer and Verma, 2001; Coffin, 1996).
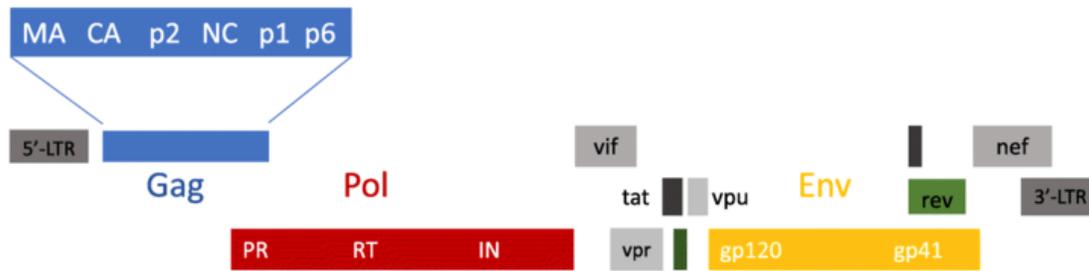
Figure 1.8 The HIV Genome. An example of a lentivirus, HIV has the major retrovirus genes (*gag, pol, env)* as well as non-structural and accessory genes specific to HIV. There are two LTR regions at the 3' and 5' end of the genome. The LTR regions are involved in virus replication, integration and expression. The *Gag* region encodes the viral proteins MA, CA and NC which are described in **figure 1.7**. It also encodes for proteins p1, p2 and p6, the former two being spacer peptides and the latter being a membrane interaction protein which acts as a docking site for several viral binding targets. The *Pol* region encodes for the viral proteins PR, RT and IN which are described in **figure 1.7**. The *env* region encodes for the proteins gp41 and gp120 also known as TM and SU and are described in **figure 1.7**. Vif (virion infectivity factor) is an essential protein for viral replication. Tat (Trans-activator of transcription) and nef (negative factor) are regulatory proteins involved in initiating viral transcription and viral replication respectively. Vpr (viral protein R) is involved in several functions including host cell cycle arrest, viral preintegration complex nuclear import and apoptosis of the host cell. Vpu (viral protein U) enhances the release of virions from host cells through their plasma membranes. The rev protein improves viral expression by inducing nuclear export of viral mRNA into the host cell cytoplasm. Image modified from Cervera *et al.*, 2019 (Gouvarchin *et* al., 2020; Cervera *et al.*, 2019; Solbak *et* al., 2013; González, 2015; Romani *et al.*, 2010; Andersen and Planelles, 2005; Chang, Liu and He, 2005; Zhang *et* al., 2000; Aiken and Trono, 1995; Arrigo *et al.*, 1989).

Lentiviruses require transport of a preintegration complex through the nucleopore of the host cell utilising its own nuclear import mechanism. This allows for the targeting of non-dividing cells which makes them an ideal candidate to use as a gene therapy vector (Vigna and Naldini, 2000). The HIV-1 life cycle can be categorised into two distinct phases, infection and replication. Infection involves the incorporation of the viral genetic information into the host genome; the extensive latent period during this is considered a unique aspect of lentiviruses. The production of regulatory products required for replication is seen early in the life cycle and is followed by structural gene expression later on, closer to the replication phase. Replication involves the final part of the life cycle involving the production of virus particles (Kay *et al.*, 2001).
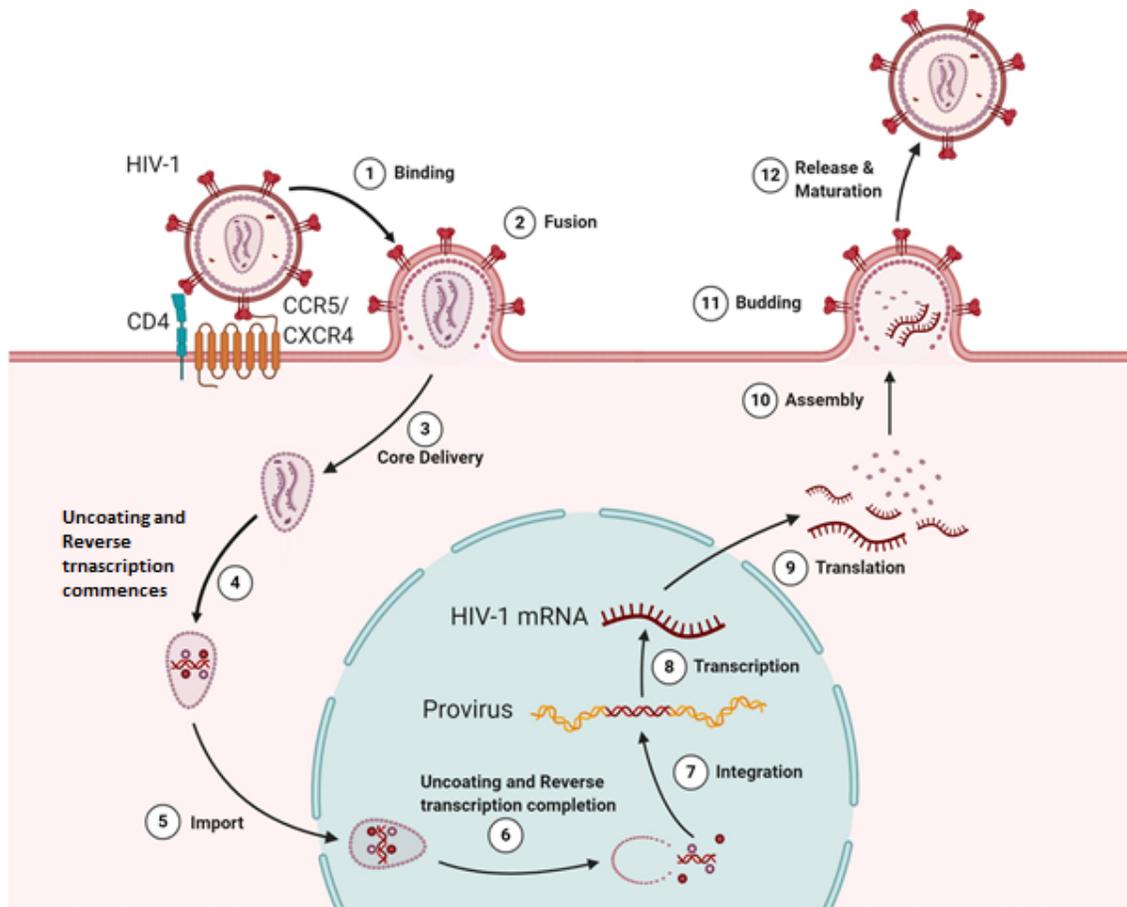
Figure 1.9 The Lentiviral Life Cycle. 1) Binding of the HIV receptor and the host cell membrane CCR5/CXCR4. **2**) Fusion of the virion transmembrane and host cell membrane allows entry of viral material into the host. **3**) Delivery to the host nucleus begins along with the uncoating process. **4**) Reverse transcription begins and the viral RNA is converted to DNA. **5**) Viral material enters the nucleus. **6**) Uncoating and Reverse transcription both complete. **7**) Viral DNA integrates into host genome. **8**) Transcription of viral DNA occurs using host mechanisms to form viral mRNA. **9**) Viral mRNA is translated to form viral proteins. **10**) Viral proteins are assembled to form a viral particle. **11**) Budding occurs to allow the virions to exit the cell and form their own virion transmembranes from the host cell membrane. **12**) The virions are released and completely mature, ready to infect more host cells. Image modified from Ramdas *et al.*, 2020.

Entry and Uncoating

The retroviral life cycle is initiated upon the entry of the virion into the cell via the env detecting a specific host cell surface receptor; the main one for HIV-1 being CD4 and co receptors CCR5 and CXCR4 (Dragic *et al.*, 1996, Choe *et al.*, 1996, Ramdas *et al.*, 2020). Upon this interaction, the surface glycoprotein (SU) will change in conformation to expose a coreceptor binding domain which will then allow the transmembrane

glycoprotein (TM) to become exposed. The hydrophobic nature of TM will cause it to fuse with the host cell membrane, allowing the virion to become engulfed (Bosch *et al.*, 1989). The virion will then uncoat by dissociating its capsid proteins while the proteins within will form the reverse transcription complex (RTC) (required for reverse transcription) and the preintegration complex (PIC) (allows entry into the nucleus for integration). The RTC contains the reverse transcriptase (RT), integrase (IN), nucleocapsid (NC), the phosphorylated matrix protein (MA) and the Vpr (Bukrinsky *et al.*, 1993; Miller, Farnet and Bushman, 1997; Gorman *et al.*, 1998).

Reverse Transcription

Reverse transcription will initiate before nuclear entry. The process utilises the tRNA (tRNAlys3) from the host cell as a primer which binds to the complementary binding site on the viral RNA strand which will initiate the DNA reverse transcription. The DNA synthesis will continue towards the 5' end of the genome encoding the U5 and R regions as the (-)ssDNA is formed. This will cause the formation of a RNA-DNA duplex while the RNA template strand is degraded by RT utilising RNase activity. The (-)ssDNA will then be transferred to the 3' end of the RNA via the homology between the 5' and 3' R sequences in a process called minus strand transfer. The DNA strand will be extended. Upon completion of extension, the complementary strand will then begin synthesis from the polypurine tract (PPT). During the formation of the strand, the viral RNA will continue to be degraded, except for the central polypurine tract (cPPT) which will act as primers for the (+)ssDNA which begins after the (-)ssDNA synthesis and will continue up to the tRNA primer, creating an region of homology at the PBS of 18 nucleotides which allows (+)ssDNA transfer. Synthesis will continue from the transferred part of the (+)ssDNA up to the cPPT where it will displace prior DNA synthesis which had been initiated from the cPPT forming a 99 nucleotide 'flap' between a central termination sequence (CTS) (involved in nuclear entry) and the cPPT. DNA synthesis continues after plus strand transfer in both directions towards the ends of the LTRs after which, the DNA can be integrated into the host genome. During the reverse transcription process, the entire complex will move toward the nucleus. Once at the nucleus, the RT will split forming the PIC and the DNA is translocated through the nuclear pore into the host nucleus. Gammaretroviruses require the host nuclear membrane to break down from mitosis as they are unable to cross the nuclear membrane. Lentiviruses are useful for gene therapy since they have a mechanism to enter non-dividing cells, making them more versatile (Rosonina *et* al. 2005; Charneau *et al.*, 1994).

Integration

This process is carried out primarily by a retroviral recombinase protein called integrase. The integrase will bind to specific attachment sites in the U3 and U5 LTRs to allow integration in two steps. The first is called 3' processing and involves the removal of the 3' terminal dinucleotide one each end of the viral cDNA resulting in a preintegration complex (PIC). This will result in 3' OH groups at the ends of the viral cDNA which act as a viral genome attachment for the host DNA (Skalka and Katz, 2005; Miller, Farnet and Bushman, 1997). In the second step, integrase will arbitrate a nucleophilic attack between the viral 3' OH groups and the phosphodiester bonds in the complementary strands of the host DNA chromosomes. The energy released by the broken phosphodiester bonds is utilised to join the 3' viral DNA to the host DNA at one end (Engelman, Mizuuchi and Craigie, 1991). The process creates a double strand break in the host DNA which is then held together by the single strand links of the viral DNA (Skalka and Katz, 2005). This will lead to around five base pairs on the 5' end of the viral DNA being unpaired thus leading to two short single stranded gaps in the host DNA flanking the integration site. These unpaired dinucleotides of the viral DNA 5' ends will then be removed and the single strand gaps between the 5' and host DNA are filled leading to the same five base pairs flanking the integrated DNA (Craigie, 2001). In order to allow the DNA to remain stable and remain heritable, post-integration repair is carried out via non homologous end joining (NHEJ) (Skalka and Katz, 2005).
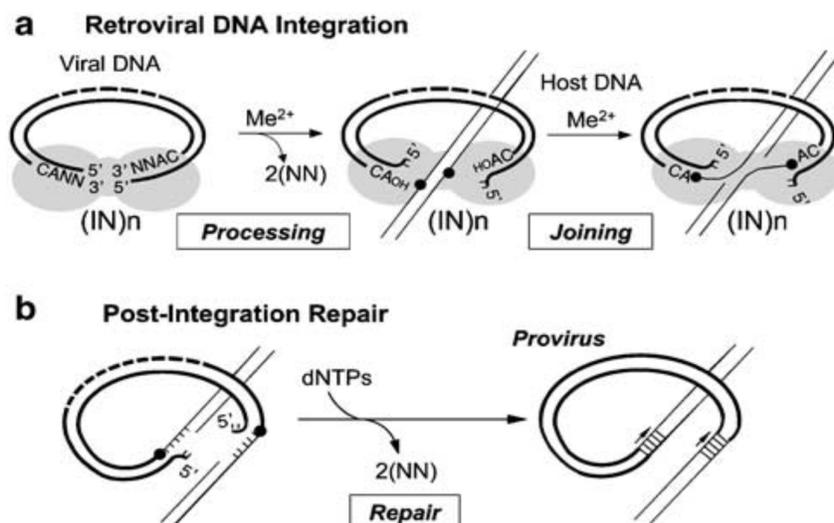
Figure 1.10 Retroviral Integration. (**a**) The ends of the viral DNA will be housed inside the integrase molecule represented as a grey dumbbell shape (the integrase is part of an integrase multimer, "(IN)n"). Me^{2+} is either Mg^{2+} or Mn^{2+} which act as a divalent metal cofactor and initiates the processing mechanism. During processing, the integrase will engulf the host strand at the integration point. This is

followed by joining (utilising more Me^{2+}) which will initiate a cleavage-ligation reaction to break the host DNA strands and ligate the 3' ends of the viral DNA to the 5' ends of the host DNA. (**b**) Postintegration repair of the strand will be carried out by host cell machinery through NHEJ. Image adapted from Skalka and Katz, 2005.

The whole process creates by products called DNA episomes which are circular viral coding regions with either one LTR (formed as a result of homologous recombination between two LTRs), or via two LTRs (created by non-homologous end joining (NHEJ)) (Cara and Reitz, 1997). This non-integrated DNA can support transcription and as such allow even defective lentiviral vectors to allow expression (Wu and Marsh, 2001; Philippe *et al.*, 2006, Yáñez-Muñoz *et al.*, 2006).

NHEJ Necessity in Integration

Prior to post-integration repair, if a host cell replication fork collides with the unrepaired viral DNA – host DNA complex, it can create DSBs with the free ends at the integration site. These unintegrated molecules can also appear as chromosomal fragments to the host DNA damage sensing repair pathways leading to self-ligation via homologous recombination and the removal of the viral DNA (which may become episomes). As such, for the successful integration of a provirus or for the use of gene therapy, NHEJ must be used in post-integration repair (Skalka and Katz, 2005; Li *et al.*, 2002).

A series of studies showed that in NHEJ-deficient cells, successful retroviral infection only occurred in 10-20% of cells. The remaining 80-90% of cells would undergo cell cycle arrest and apoptosis in the presence of an active integrase molecule. (Daniel, Katz and Skalka, 1999; Daniel *et al.*, 2001; Daniel *et al.*, 2004). Thus, without NHEJ capabilities, the DNA integration process is most likely detected by the host cells as DNA damage. So, a lack of NHEJ would prevent the post-integration repair necessary for successful integration leading to cell cycle arrest and apoptosis. It is possible the cells can still undergo repair through homologous recombination (HR) at a lower rate, which may account for the 10-20% of successful retroviral infections (Skalka and Katz, 2005; Li *et al.*, 2001).
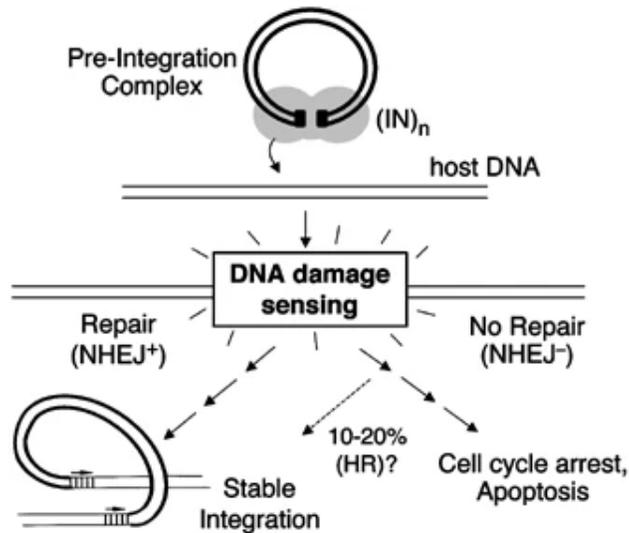
Figure 1.11 DNA Damage Sense. DSB formation during viral integration will be perceived as DNA damage by the host cell. NHEJ competent cells will repair the damage leading to stable integration. Only 10-20% of cells lacking NHEJ will be able to achieve stable integration, possibly via homologous recombination (HR), while the majority of cells will undergo cell cycle arrest and apoptosis. Thus it can be understood that NHEJ is crucial for successful retroviral infection. Image adapted from Skalka and Katz, 2005 (Skalka and Katz, 2005; Li *et al.*, 2001).

The impairment of NHEJ can lead to a multitude of conditions related to cell cycle checkpoints such as Artemis-dependent SCID (Moshous *et al.*, 2001; Kerzendorfer and O'Drriscoll, 2009). Thus, the induction of an increased number of double strand breaks in DNA in NHEJ impaired cells could potentially lead to serious effects in patients being treated with gene therapy.

Transcription

The provirus, once formed after integration, will utilise the host cell machinery to transcribe and translate its genetic information. The *cis*-acting regulatory sequences in the LTR and the *trans*-acting proteins made by the host and the virus interact in order to allow for transcription. The 5' end of all mRNA transcripts will recruit host cell elements including cyclin T and Cdk9 upon the viral transcription activator (Tat) protein attaching to the stem loop structure called the trans-activation response (TAR) element at the 5' end of these transcripts. This will allow for transcription to occur at a much higher rate (close to 100 times more) than the normal rate acquired by the HIV-1 promoter mediated RNA polymerase complex (Romano and Giordano, 2008). Unspliced mRNA will be bound to by Rev, at the Rev responsive element (RRE) allowing this complex to eject

from the nucleus where the Rev will stabilise the viral transcripts (Roebuck and Saifuddin, 1999; Pollard and Malim, 1998, Schneider *et al.*, 1997).

Translation

The Gag and the Gag-Pro-Pol polyproteins are translated from unspliced viral mRNA at a ratio of 20:1, after which they associate to the cellular membrane where Gag multimerisation happens as the viral genome is attached to the nucleocapsid (NC) via the encapsidation signal present on unspliced mRNA leading to packaging. The *env* gene is transcribed as a dull mRNA and translated to the viral glycoproteins gp120 and gp41 and inserted into the rough endoplasmic reticulum via the signal recognition particle (SRP) following which it will be recruited into the virion (Perez *et al.*, 1987).

Virion Assembly and Budding

The virions are then packaged at the cell membrane with host proteins and released from the cell as immature and non-infectious particles. The polyproteins in the virions will then be cleaved by the protease (P) to form the IN, RT and the structural proteins, MA, CA and NC. This will allow the immature spherical gag to be condensed into the mature cone shaped core, forming a mature virion. (Briggs *et al.*, 2003).

1.2.2.3 Lentiviral Vectors

Viral vectors were first used for gene delivery in the early 1980's utilising the Moloney Murine Leukaemia Virus (Mo-MLV). In order to utilise vectors, it is necessary to separate the viral particle formation proteins, the infection proteins and the *cis* acing sequences into different plasmids (Mann *et al.*, 1983). Lentiviruses are characterised by their long period between the infection and the onset of disease, sometimes lasting months or years, as well as their ability to infect non-dividing cells (Buchholz *et al.*, 2009). The infection phase of the lentivirus life cycle, which involves the genome being introduced to the host cell, is the aspect which allows these lentiviruses to be used as vectors and is referred to as gene transduction or single cycle infection (Durand and Cimarelli, 2011).

Other aspects which allow lentiviruses to be used as vectors included the fact that the reverse transcription and integration processes are carried out inside the virion itself, allowing straightforward access to the viral genes which must be altered before use as a vector. They also have a large capacity of 8-10 kb aside from the *cis*-acting sequences

(Zufferey *et al.*, 1998) and have reduced *in vivo* immunogenic reactions as compared to other viral gene therapy vectors (Chirmule *et al.*, 1999). The integration of material into the host cell genome allows subsequent generations to have the transgene.

Early HIV-1 Vectors

The very first lentiviral vectors were able to replicate and were used as a tool to detect viral replication *in vitro* and *in vivo* to screen for anti-HIV-1 drugs. Subsequently, to improve safety of these HIV vectors, they were modified to separate the viral sequences required for packaging and those needed to form the viral proteins with the first prototypes separating the information into two plasmids (Helseth *et al.*, 1990, Page *et al.*, 1990). The first plasmid contained the HIV-1 proviral DNA without the *env* gene while the second only expressed the *env*. The viruses would undergo one cycle of infection since they lacked the *env* and instead had the transgenes inserted instead of the nef or env; expression controlled by the 5' LTR (Page *et al.*, 1990, Landau *et al.*, 1991, Helseth *et al.*, 1990).

Other HIV-1 vectors were formed which carried the *cis*-acting elements for genome packaging, reverse transcription and integration while lacking the viral proteins with a heterologous internal promoter diving the expression (Richardson, Child and Lever, 1993).

CD4 Specificity and VSV-G

Early HIV vectors required human cells with CD4 to be able to express them as this was the primary HIV-1 Env receptor. However, a pseudotyped murine leukaemia virus based retroviral vector (MLV) was produced by replacing the Env glycoprotein with the vesicular stomatitis virus-G (VSV-G) viral attachment protein. The advantages of these vectors was that they were more stable and allowed higher concentration titre production via ultracentrifugation (Burns *et al.*, 1993) and the fact that VSV-G could attach to ubiquitous receptors allowing a larger variety of host cells to be infected, including non-mammalian (Coil and Miller, 2004). This allowed Akkina *et al.* to show the VSV-G HIV-1 vector was able to infect CD34+ haematopoietic stem cells (Akkina *et al.*, 1996).

HIV-1-Based Lentiviral Vectors with Three Plasmids

To further improve safety, three plasmids were used in the first generation HIV-1 vectors to prevent the formation of replication competent lentiviruses (RCLs) (Naldini *et al.*, 1996). The three plasmids contained the packaging construct, the *env* plasmid (coding for viral glycoproteins) and the transfer vector genome construct. The packaging construct contained the gag, pol and regulatory proteins necessary to make a mammalian promoter to allow viral particle generation. The *env* plasmid coded for the viral glycoprotein, such as VSV-G, to allow the vectors to bind to receptors on the host cell. These two plasmids were formed lacking the packaging signal and the LTRs to prevent the formation of RCLs. The transfer vector contained the *cis*-acting elements required for packaging, reverse transcription and integration, while lacking HIV proteins. The transfer genome required the use of an internal promoter due to the lack of the transactivator (Tat) not being encoded by it. Safety was improved with this system by the necessity of at least two recombination events occurring to produce RCLs (Durand and Cimarelli, 2011).

Second Generation Lentiviral Vectors

To improve safety, accessory genes were modified to attempt to improve these vectors. By removing the accessory HIV-1 genes, *Vif*, *Vfu, Vpr* and *Nef*, certain human lymphoid cells could still be infected without altering viral replication. These are necessary for HIV-1 infection in primary cells or *in vivo* (Fouchier *et al.*, 1996).

Nef degrades host proteins such as CD4 and MHC class I to increase virus production and to evade the immune system. Vif and Vpu are involved in the inactivation of host antiviral factor, apolipoprotein B mRNA editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) (Harris *et al.*, 2003) and the antiviral factor tetherin, respectively (Neil *et al.*, 2008, Sakuma, Barry and Ikeda, 2012). While important for viral pathogenicity, they could be deleted in second generation lentiviral vectors. A further replacement of the Env with VSV-G leads to the virus only containing four of the nine HIV genes: *gag*, *pol*, *tat* and *rev* (Zufferey *et al.*, 1997).

Self-Inactivating (SIN) Vectors with a U3 Deletion

If replication-competent recombinant lentiviruses are produced, or if wild-type lentiviral infection occurs following vector transduction, then the virus can spread the transduction of material to non-target cells. LTRs also contain enhancers which allow for host

transcription factor binding, and thus if integration occurs near proto-oncogenes, these enhancers could cause oncogenesis. These aspects have required the production of self-inactivating (SIN) lentivectors. The two LTR regions flanking the vector genome in standard vectors contain the U3 which acts as a viral enhancer/promoter while the R acts as the polyadenylation signal. Generally, the U3 and U5 regions are not present in the proviral mRNA, with the R region capping it at both ends instead. During reverse transcription, the U3 in the 3' LTR is copied to the 5' LTR thus leading to duplication of the LTR. If part of the U3 in the 3' LTR is removed, the duplication will result in the same deletion being transferred to the 5' LTR enhancer/promoter region thus preventing the transcription of viral genomes which could be packaged into replication-competent lentiviruses (RCLs) (Yu *et al.*, 1986). The SIN method was utilised with HIV vectors by deleting the 3' LTR sequences such as the TATA-box, specificity protein 1 (Sp1), and nuclear factor of activated T-cells (NFAT) binding sites (Miyoshi *et al.*, 1998, Zufferey *et al.*, 1998, Iwakuma, Cui and Chang, 1999). This reduced the chance of spontaneous production of RCLs, activation of proto-oncogenes by the LTRs, wild type virus infections utilising integrated vector proviruses and the transcriptional interference and suppression of genes adjacent to the vector provirus by the LTRs.

Third Generation Four Plasmid Tat-Independent Vectors

Tat and Rev are necessary for HIV-1 replication and thus cannot simply be deleted. They are required for transcription of the provirus and the nuclear export of transcripts containing the introns (Laspia *et al.*, 1989, Terwilliger *et al.*, 1988). To improve safety, the Rev was provided from a separate plasmid and producing a Tat-independent vector by replacing the U3 promoter region of the 5' LTR with viral promoters from cytomegalovirus (CMV) or the Rous Sarcoma Virus (RSV) (Dull *et al.*, 1998, Kim *et al.*, 1998). The four plasmids utilised were the packaging construct, the Rev plasmid, the Env or VSV-G plasmid and the transgene plasmid with a strong promoter.

The SIN method was also used with this to produce a vector with only three of the nine HIV genes improving the safety of the vector since three recombination events would be required to make a seemingly replication competent HIV-1, and even if these occurred, no LTRs, Tat or accessory proteins would be present (Escarpe *et al.*, 2003).

Cis-acting cPPT for Increased Vector Transduction Efficiency

The plus strand DNA is synthesised from the PPT and the cPPT leading to the formation of the triple helical central DNA 'flap' (Charneau *et al.*, 1992). This flap is thought to improve HIV-1 proviral DNA nuclear import (Zennou *et al.*, 2000). The introduction of the cPPT in HIV vectors improves transduction efficiency *in vitro* and *in vivo* (Zennou *et al.*, 2001, Van Maele *et al.*, 2003, Demaison *et al.*, 2002).

WPRE Introduction for Increased Transgene Expression

WPRE (WHV [woodchuck hepatitis virus] post-transcriptional regulatory element) is a *cis*-acting element which improves expression of lentiviral vectors by improving the amount of unspliced RNA in the nucleus and the cytoplasm (Donello *et al.*, 1998, Zufferey *et al.*, 1999). Adding WPRE to lentivectors improves transgene expression in target cells (Zufferey *et al.*, 1999, Brun *et al.*, 2003). However, WPRE contains the WHV X gene which is thought to be involved in animal liver cancer (Kingsman, Mitrophanous, and Olsen, 2005). To overcome this problem, the X gene open reading frame can be mutated to improve the safety of the vector (Zanta-Boussif *et al.*, 2009; Rémy *et al.*, 2009).

SFFV LTR Promoters

Promoters are regions upstream of genes which are *cis*-acting regulatory regions which control transcription. They work by providing binding sites for RNA polymerase II (transcribes DNA to mRNA) and transcription factors. They can work in conjunction with regulatory elements like enhancers and silencers to control the level of a gene's expression (Frecha *et al.*, 2008). *In vivo* gene therapy functions by expression the transgene at the required amount, at the correct developmental stage, without inducing toxicity and without being targeted by the immune system. Without controlling the expression, depending on the condition being treated, over or under expression of the transgene can have undesired effects. Naturally existing promoters have been studied to see the possibility of utilising them in order to improve transgene expression (Frecha *et al.*, 2008).

One promoter which has been studied and utilised is the Spleen focus forming virus LTR (SFFV) promoter. The U3 region in the LTR contains the enhancer/promoter with *cis*-acting elements. The 5' end is flanked by a binding site for the transcription factor, Sp1, and the 3' end is flanked by the Friend Virus factor c (FVc) (Baum *et al.*, 1997). The

promoter contains a region allowing the CCAAT/enhancer binding protein to attach which allows for liver specific transgene expression. It also contains the core binding factor which regulates retroviral gene expression (Baum *et al.*, 1998).

Utilising this gene in a lentivector showed high transgene expression in haematopoietic cells (Tsuji *et al.*, 2000), spermatogenic cells (Danno *et al.*, 1999) and hepatocytes *in vivo (*Yamaguchi *et al.*, 2003*)* indicating the viability and use of these promoters to enhance transgene expression.

1.2.3 DNA DSB Repair

Double strand breaks (DSBs) are repaired by all eukaryotic cells via two main mechanisms, homologous recombination (HR) and non-homologous end joining (NHEJ). HR will form an exact copy of the DNA region and use it as a template to resynthesise the DSB. It involves the use of a sister chromatid and is only done within the S phase of the cell cycle. If the replication fork is interrupted, the HR mechanism will reinitiate the replication thus providing a very accurate repair mechanism. NHEJ is less accurate but provides a quicker method by simply refusing the DSB. The drawback is the fact that the removal or addition of some nucleotides may be necessary, possibly leading to errors. The NHEJ repair mechanism is done throughout the cell cycle, but primarily at the G1 phase (Mohiuddin and Kang, 2019; Shrivastav, De Haro and Nickoloff, 2008; Jeggo and Lobrich, 2007; Essers *et al.*, 2000).

1.2.3.1 DNA Repair by Homologous Recombination (HR)

An MRN complex consisting of Mre11p, Rad51p and NBS1 will process the broken ends in the 5' to 3' direction (Shin *et al.*, 2004). The replication protein A (RPA) will bind the 3' ends of the single stranded DNA. RPA will then be phosphorylated and detach allowing Rad52 to bind, allowing Rad51 to bind next. The homologous sequence which is complementary to the DSB region will be bound by the Rad51 to form a nucleoprotein. Rad51 will detach and normal base pairing by DNA polymerase will occur, extending the ssDNA strand based on the complementary DNA sequence. Synthesis-dependent strand annealing occurs at this point where the extended ssDNA is annealed with the non-invading DNA strand to produce a double Holiday junction which is then resolved by crossover or non-crossover recombination. DNA polymerase and ligase will then repair the nicks and gaps formed during the ligation of the two ends (Shrivastav, De Haro and Nickoloff, 2008; Shin *et al.*, 2004; van Gent, Hoeijmakers and Kanaar, 2001).

1.2.3.2 DNA Repair by Non-Homologous End Joining (NHEJ)

The Ku70 and Ku80 subunits forming the Ku protein detect the overhanging ends of the DSB and will attach to the DNA to prevent further damage and to allow repair proteins including DNA-dependent kinase catalytic subunit (DNA-PKcs) access (DS, 2005). The non-ligatable end groups will then be removed utilising a variation of enzymes depending on the properties of the break such as DNA polymerases, MRN complexes, RPA and WRN. The XLF (XRCC4-like Factor) will then induce the XRCC4/DNA ligase IV to join the ends (Summers *et al.*, 2011; Shrivastav, De Haro and Nickoloff, 2008).

NHEJ is inaccurate due to the trimming which occurs at the ends. In cases where two breaks occur, the ends can become mixed up and cause translocation of materials leading to conditions such as Burkitt's lymphoma which is caused by moving inactive c-myc genes to a promoter heavy region causing uncontrolled cell growth due to over expression of the gene (Rowh *et al.*, 2011).
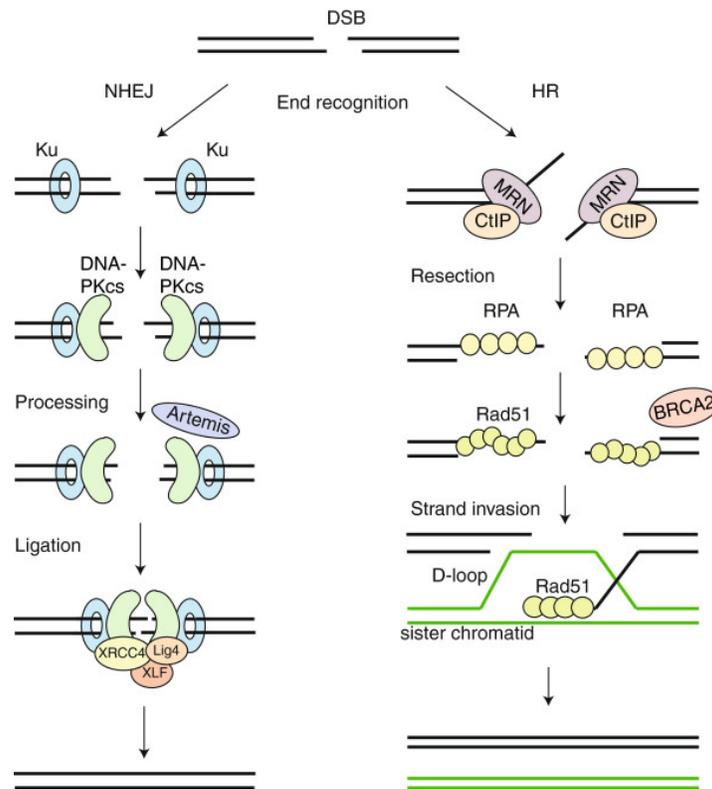
Figure 1.12 Differences between NHEJ and HR. NHEJ: DSBs are first detected by the Ku 70/80 heterodimers which will then direct DNA-PKcs to the DSB to allow kinase activity and processing of the DSB. The nuclease Artemis will remove any nucleotides that prevent complementation between the two strands. The XFL-XRCC4-DNA Ligase IV complex will then anneal the broken ends of the DSB together. **HR:** The MRN-CtIP complexes will detect and bind to the ends of the DSB and begin to form

ssDNA. RPA will coat the ssDNA portion which will them be replaced by Rad51 which allows ssDNA to invade the sister chromatid where the ssDNA will align with the complementary sequence via the formation of a D-loop. Homologous repair can then occur (Mohiuddin and Kang, 2019). Image adapted from Brandsma and Gent, 2012.

1.2.3.3 Determining the Presence of DSBs using γH2AX

H2AX is a histone protein found in chromatin and at DNA DSB sites (McKinnon and Caldecott, 2007; Rogakou *et al.*, 1998). It contains a serine residue which is readily phosphorylated into γH2AX by the protein kinase family ATM in the case of DSBs (Bassing *et al.*, 2003; Rogakou *et al.*, 1998). It does not diffuse into the cell freely, suggesting the phosphorylation is important for repair and not for cell cycle arrest (Zgheib *et al.*, 2005). It also causes the accumulation of DNA damage repair factors (DDR) at DSBs. Thus, detection of γH2AX via the use of antibody directed fluorescence is a very straightforward method of detecting DSBs.

1.3 Genotoxicity

Genotoxicity is a property of a chemical agent to induce an uncontrolled change in the genetic material of a cell and leads to a mutagenic change in phenotype such as cancer (Ramezani, Hawley and Hawley, 2008). Retroviral vectors directly alter the genes of a host cell and as such must not be genotoxic as a side effect.

One key feature of the retroviral vector which can cause genotoxicity is integration. As described earlier, the process involves the formation of double strand breaks in a semi-random manner within open chromatin regions which are actively being transcribed (Albanese *et al.*, 2008; Baum *et al.*, 2004; Mitchell *et al.*, 2004; Wu *et al.*, 2003; Schröder *et al.*, 2002). The semi-random nature of the retroviral integration process can disrupt the genome of the host cell and cause certain genes to become active or inactive resulting in genotoxicity (Nienhuis, Dunbar and Sorrentino, 2006). Wild-type viruses have been known to cause insertional mutagenesis leading to tumorigenesis by either altering the expression of an oncogene in the host genome by integrating near it or by carrying part of an oncogene which is then expressed uncontrollably. The ability of retroviruses to cause these mutagenic changes has even been used to study cellular mechanisms by altering the normal behaviour of the cell (Varmus, 1982; King *et al.*, 1985).

There have not been many studies on the genotoxic effects of retroviral vectors as the chances of this happening has been considered remote at around 10^{-7} for a haploid locus (Stocking *et al.*, 1993).

1.3.1 Insertional Mutagenesis

While most DNA sequences will allow for retroviral integration, once a site has used as an integration site, it will induce the formation of insertion site "hot spots" as seen in cases where DNA is placed into nucleosomes *in vitro*. This is thought to be caused by the insertion sites being distorted when DNA wraps around nucleosomes due to a prior insertion; this, in turn, will lead to different areas of the host DNA becoming accessible for integration (Pruss *et al.*, 1994). Retroviral insertion site analysis can be done using molecular methods which can retrieve the DNA at the spot where integration has occurred allowing for the analysis of several thousand integrations (Mitchell *et al.*, 2004; Schmidt *et al.*, 2002; Schröder *et al.*, 2002). Analyses done with MLVs have shown that open chromatin is preferred as seen with the correlation between DNase I-hypersensitive sites and integration spots (Rohdewohld *et al.*, 1987; Vijaya, Steffen and Robinson, 1986). Thus, because integration leads to more sites forming open chromatin, an integration will lead to a cascade of integrations as more open chromatin becomes available for the viruses.

A comparison of lentiviral vectors such as HIV-1 and EIAV to MLVs showed that the lentiviral vectors were less genotoxic, preferring insertion in gene transcription units. MLVs preferred insertion in areas near promoters and genes involved in oncogenesis (Montini *et al.*, 2009; Cherepanov, 2007; Bushman *et al.*, 2005; Schröder *et al.*, 2002). Studies have shown that viruses can also deform the proteins produced by genes due to aberrant splicing of the viral and host DNA. Should these affected genes be involved in tumour suppression and are inactivated, or involved in proto-oncogenesis and are activated, this can lead to uncontrolled cell division and tumour formation (Modlich and Baum, 2009; Uren *et al.*, 2005; Baum *et al.*, 2004).

1.3.2 Insertional Mutagenesis History

1.3.2.1 X-SCID

X-SCID is a monogenic X-linked disorder caused by mutations in the cytokine IL2 receptor gamma-chain (IL2RG) (Howe *et al.*, 2008; Thomas, Ehrhardt and Kay, 2003;

Cavazzana-Calvo *et al.*, 2000). IL2RG forms part of a cell surface receptor in developing immune cells for cytokines to allow for growth. The absence of the subunit will prevent the development of mature T lymphocytes, thus causing an additional effect of B lymphocytes to be unable to produce antibodies (Hacein-Bey-Abina *et al.*, 2008; Thrasher *et al.*, 2006). The condition will cause a very low amount of T, B and natural killer (NK) cells thus putting very young patients at fatal risk to recurrent infections as they have a low amount of immunoglobulin early in life (Thrasher *et al.*, 2006).

One treatment for this condition is via bone marrow transplant from HLA-matched donors. However, it is difficult to avoid immune rejection due to mismatch. Gene therapy was sought as an alternative method. In the year 2000, gene therapy was used to restore T, B and NK cells in gamma-C deficient mice using a retroviral vector with the gamma-C gene (Lo *et al.*, 1999). The major concern at the time was the lack of control over integration sites and cancer development was under consideration (Check, 2002).

As mentioned in the history of gene therapy section, French and British trials altered haematopoietic stem cells from children suffering from X-SCID and altered *ex vivo* with an MLV virus with the *gamma-C receptor* gene. These cells were then reintroduced back into the patients (Cavazzana-Calvo *et al.*, 2000; Hacein-Bey-Abina *et al.*, 2003; Gaspar *et al.*, 2004). T and NK cell function was seen to be restored after several months but five patients developed leukaemia three to five years later (Hacein-Bey-Abina *et al.*, 2010; Qasim, Gaspar and Thrasher, 2009).

LAM-PCR was used to try to identify if insertional mutagenesis was the cause of the leukaemias. LAM-PCR allows the capture of DNA sequences adjacent to an insertion site allowing for the analysis of those sequences. This showed that four of the five patients with leukaemia had integrations near the LIM domain only 2 (LMO2) proto-oncogene. Elevated LMO2 expression had caused uncontrolled proliferation of mature T cells (Hacein-Bey-Abina *et al.*, 2010; Qasim, Gaspar and Thrasher, 2009; Nam and Rabbitts, 2006; Hacein-Bey-Abina *et al.*, 2003). It remained unclear why only 4 of 11 children had this integration; however, a theory pointed towards gene expression causing insertion site selection (Kaiser, 2003; Coutelle *et al.*, 2003). Transplanting bone marrow expansions from the transduced bone marrow cells into mice caused them to develop leukaemia (Li *et al.*, 2002). While the transgene itself was thought to cause growth promotion, it was suggested that a combination of the vector product along with the actual MLV integration disrupted a proto-oncogene (Baum *et al.*, 2003). It was also

seen that the use of a high multiplicity of infection (MOI) would induce leukaemia in mice and several insertion sites were found near proto-oncogenes (Modlich *et al.*, 2005).

1.3.2.2 Chronic granulomatous (X-CGD) trial

X-linked Chronic Granulomatous Disease (X-CGD) is caused by mutations in the *gp91phox* (*CYBB*) genes in two out of three cases of the condition. Other causes are abnormalities in the *p22phox (CYBA)* and *P67phox (NCF2)* genes. The *gp91phox* will code for the NADPH oxidase enzyme which is involved in superoxide production from oxygen and NADPH (Kang and Malech, 2009; Seger, 2008). The lack of NADPH activity causes neutrophils, monocytes and other phagocytic cells being unable to produce the reactive oxygen species needed to destroy bacteria leading to frequent, possibly fatal, infections (Stein *et al.*, 2010; Malech *et al.*, 1997; Björgvinsdottir *et al.*, 1997). X-CGD, like X-SCID, is also treated with bone marrow transplants and thus faces similar issues with immune rejection due to mismatch. The genes responsible for X-CGD are known, and thus gene therapy was attempted to cure the condition.

In 2001, retroviral transduction was successfully utilised on rats with a *gp91phox* mutation following which the rats were able to produce reactive oxygen radicals (Dinauer *et al.*, 2001). Following this, in 2008, bone marrow stem cells from mice with the transduced *MT-gp91phox* gene and the mice were monitored for toxicity. While white blood cell counts increased, there was no toxicity (Lee *et al.*, 2008). Thus, it was seen that clonal imbalance was not caused by gamma retroviruses unlike the X-SCID trial.

In 2004, *gp91phox* gene transduced CD34[+] blood stem cells were infused in X-CGD patients. The results were initially positive. Follow up analysis showed clonal dominance of haematopoietic clones with insertions in specific gene loci (Gaspar *et al.*, 2004). Upon analysing the gene inserts, it was thought that the gene-activating and/or supressing effects of the integrated viral vector led to the clonal dominance seen (Fehse and Roeder, 2008; Ott *et al.*, 2006). This result, among others, led to the belief that a low copy number of vectors will reduce the likelihood of oncogenesis (Ramezani, Hawley and Hawley, 2008). Other factors seen to affect clonal dominance were seen to be vector configuration, the specific transgene, host cell proliferation status at the time of transduction, synergy of the vector transcription status, and mutagenesis potential (Nowrouzi *et al.*, 2013; Baum *et al.*, 2006). Thus, we see that vector insertion and genotoxicity needs to be further explored to improve the safety of gene therapy vectors.

1.3.3 Models of Genotoxicity

To avoid vector-associated adverse side effects of gene therapy, it is important to understand how exactly these vectors may affect the host in detail. Thus, *in vitro* and *in vivo* models have been established to try and determine the genotoxic effects that can occur due to the use of these vectors.

1.3.3.1 *In Vitro* Models of Genotoxicity

An adaptation of the *HPRT* assay has been used to determine vector-mediated genotoxicity. The assay utilises the *HPRT* gene encoding for hypoxanthine-guanine phoshoribosyltransferase (HPRT) protein, involved in conversion of purines into monophosphates that are toxic to cells. The V79 Chinese Hamster cell line is used since it is a male cell line and thus only has one copy of the *HPRT* gene on the X chromosome (Zhang *et al.*, 1994). HAT treatment can be used to remove all cells with *HPRT*-mutants ensuring a complete population of cells only containing the functional *HPRT* gene. These cells can then be exposed to a virus which will then cause insertional mutagenesis. All new *HPRT*- mutants following this will likely to have be caused by viral induction. The mutants can be selected using 6-thioguanine or 8-azaguanine, which will be converted by non-mutated *HPRT* to toxic monophosphates leaving only cells with mutated *HPRT* to survive.

This method was used to determine if retroviral insertional mutagenesis causes loss of *HPRT* activity (Themis *et al.*, 2003). While prior predictions placed a single provirus insertion in a haploid locus of the mammalian genome would occur at one cells per 10^6 cells (Goff, 1987), experiments showed that this mutagenesis occurred at a rate of one per 10^8 provirus insertions (King *et al.*, 1985). It was eventually seen that *HPRT* mutagenesis can occur at around one in 3.6×10^6 if a high MOI is used. A 2.3-fold increase in mutagenesis was also seen if multiple provirus insertions per host genome occurred (Themis *et al.*, 2003).

The relationship between vector copy number and gene transfer efficiency was done using K562 leukaemia cells and primary CD34+ cells by Kustikova *et al.* in 2003. It was seen that insertional mutagenesis frequency was linked to vector copy number. A single cell transduced with a single vector showed up to 30% gene transfer efficiency. 3 vectors per cell increased efficiency to 60% and proportionally, 9 per cell led to a 90% transfer efficiency (Kustikova *et al.*, 2003).

Clonal dominance leading to malignancy was investigated by growing bone marrow stem cells *ex vivo* and then inserting them into primary and then secondary mice. A database called the insertional dominance database (IDDb) was generated which showed insertion sites relating to malignancy (Kustikova *et al.*, 2007). The purification of haematopoietic stem cells was then investigated to see if it had an effect on insertional mutagenesis. While this was not seen to reduce the genotoxic effect of retroviral transduction, the reduced number of haematopoietic stem cells did lower the risk of genotoxicity (Kustikova, Modlich and Fehse, 2009). It was also seen that compared to gamma-retroviral vectors, lentiviruses reduced the risk of clonal imbalance of provirus insertion into proto-oncogenes (Kustikova, Modlich and Fehse, 2009).

Other experiments showed that an increase of vector dose was seen to increase insertional mutagenesis (Modlich *et al.*, 2005). Bone marrow cells from C57BL/6J mice were treated with retroviral vectors with the *MDR1* gene *ex vivo*. The cells were then returned and the development of leukaemia appeared to be associated with a high MOI infection. The cells were seen to have a high number of proto-oncogene and other signal gene insertions (Modlich *et al.*, 2005). Other factors also contributed to genotoxicity, including the architecture of the virus. The relocation of strong enhancer regions from the LTR region showed a reduction in genotoxicity. A correlation was also seen between *Evil* gene insertion as well as the MOI used (Josephson and Abshire, 2006; Modlich *et al.*, 2006; Wang *et al.*, 1999). Thus, the principles observed could be applied to other viral vectors in order to reduce genotoxicity.

1.3.3.2 *In Vivo* Models of Genotoxicity

Genotoxicity studies primarily use the mouse model because of their small stature, rapid reproductive capabilities while still having biological and behavioural similarities to humans. Larger models have also been used, including sheep, pigs, monkeys and dogs (Amsterdam *et al.*, 1999; Tarantal *et al.*, 2001). *In vivo* methods can make use of the hepatocellular carcinoma (HCC) phenotype, a valuable marker to determine genes involved in liver tumorigenesis. The model avoids the use of cell engraftment and proliferation (as is required in certain *ex vivo* methods). The method also allows the identification of molecular pathways which lead to immortalisation and malignant progression which differ from leukaemia.

In utero studies have been conducted in both large and small animal models (Tarantal *et al.*, 2005; Themis *et al.*, 2005; Walsh, 1999). A mouse model was developed based on

gene therapy before birth. The model exhibited multiple genetic disorders in the foetus before birth and *in utero* gene therapy is necessary to treat conditions before birth by targeting stem cells, gene delivery which avoids immune rejection to the vector and transgene product, tolerance to the vector and permanent correction (Coutelle *et al.*, 2005).

Human factor IX (FIX) deficiency was corrected in FIX mice utilising HIV-1 lentiviral vector which allowed FIX levels to increase in the plasma without causing an immune response (Waddington *et al.*, 2004).

Viral insertion sites are believed to target genes that are actively dividing, as is the case in rapidly dividing foetal cells, and as such, the risk of insertional mutagenesis was investigated. A foetal mouse model was used to observe liver cancers in mice treated *in utero* (Themis *et al.*, 2005; Nowrouzi *et al.*, 2013). Both primate and non-primate vectors, HIV and EIAV, were used on the model to increase FIX production. Both corrected the haemophilia B knockout mice. However, the EIAV treated mice developed hepatocellular carcinoma (HCC) in 8 out of 10 cases (Themis, 2005). The high level of FIX expression allowed the prediction of these tumours forming. DNA from these tumours showed 1 to 10 integrated proviruses per genome. LAM-PCR allowed the mapping of EIAV insertion sites in relation to genes in the mouse genome. 56% of these sites were either oncogenes or associated with oncogenes. 99% were reduced in expression indicative of insertional mutagenesis, 11 of which tagged in the retroviral tagged cancer gene database (RTCGD) as known genes involved in tumorigenesis in mice following retroviral infection (Themis *et al.*, 2005; Akagi *et al.*, 2004). The primate HIV vector was not seen to cause tumour development and as such could be used for prenatal gene therapy (Waddington *et al.*, 2004).

A foetal model was used to show insertional site preference in actively transcribing genes and the different insertion site profiles between primate and non-primate vectors (Nowrouzi *et al.*, 2013).

1.4 HERVs

Human endogenous retroviruses (HERVs) make up approximately 8% of the entire human genome and are mainly dormant endogenous retroviral sequences. HERVs are thought to be partial to whole sequences of viruses which had infected humans ancestrally over millions of years; and over time have evolutionarily lost their infectious nature due to mutations and internal recombination (Steinhuber *et* al., 1995; Bushman *et*

al., 2005; Feschotte and Gilbert, 2012; Stoye, 2012; Dewannieux and Heidmann, 2013; Contreras-Galindo *et al*., 2015; Wildschutte *et* al., 2016; Gifford *et al*. 2018; Khalfallah and Genge, 2021). The HERVs are not seen to have fixed locations. This has been shown through the analysis of HERV-Ks which have been found in pieces or even in their entirety at several points within the human genome (Wildschutte *et* al., 2016). One KERV-K, known as HERV-K113, has been seen with its entire coding capacity including the *gag, pol* and *env* genes along with the 5' and 3' LTR regions (van der Kuyl, 2012; Beimford *et al*., 2008).

HERV LTR regions have been seen to inhibit or activate promoter and enhancer activity around their locus in the human genome (Jern and Coffin, 2008). HERVs are also able to cause the formation of novel splice locations and induce their activation (Cohen, Lock and Mager, 2009).

HERVs have shown to influence the immune system, causing aggravating effects, such as inflammation and symptoms related to autoimmune disorders, but also displaying immunosuppressive properties (Grandi and Tramontano, 2018). HERVs have also been shown to cause neurotoxic effects by over activating certain genes such as *NTRK3* (Nair *et al*., 2022). They have also been used as markers for neurodegenerative diseases such as multiple and lateral sclerosis due to the significant correlation seen between the diseases and the up-regulation of HERVs (Dolei *et* al., 2019; Ramussen *et al*., 1997). While thought to be dormant, factors such as irradiation and internal cell signalling have been seen to induce HERV expression (Baier, Morell and O'Carrol, 2022; Min *et al*., 2022). This up regulation is thought to be a major part immune responses seen in lupus erythematosus patients following radiation (Min *et al*., 2022). HERVs have also been seen to be transcribed in cancer cell lines, present in cells with pancreatic cancer, breast cancer, prostate cancer and melanomas (Li *et* al., 2019; Agoni, Luha and Lenz. 2013; Büscher, K, 2005; Sauter *et al*., 1995, Frank *et al*. 2008; Ishida *et al*., 2008).

While HERVs can undergo translation and translation, they are thought not to be able to transfer between cells by themselves. Experiments have shown that a specific HERV known as HERV-K has been shown to demonstrate expression into complete viral particles. This was seen when one of the provirus variant, HERV-K113, was packaged into baculovirus expression vectors, which were then able to synthesise complete HERV-K113 viral particles (Boller *et al*., 2007).

HERV-K113 expression has been seen to increase in cells infected by HIV-1. This expression has even reached levels where HERV-K113 RNA was detected in the

plasma of patients (Contreras-Galindo, *et al.*, 2007; Contreras-Galindo, *et al.*, 2006). While HIV and HERVs have been shown to have very little complementation activity (van der Kuyl, 2012; Ogata *et al*, 1999), HIV-1 based packaging cell lines were able to package HERV transcripts (Zeilfelder *et al.*, 2012; Rulli, *et* al., 2007; Sakai *et al.*, 1990).

The HERV known as HERV-K (HML2), shortened to HK2, has been found to be intact enough to form viral particles. These particles, while thought to be non-infectious, have been seen to be packaged and have been transmitted to other cells. Inside those cells, the HK2 RNA was reverse-transcribed, but unable to integrate into the host cell. No replication-competence has been seen (Contreras-Galindo *et al.*, 2015).

Cell lines which have been used to produce RV and LV vectors have been seen to package HERVs. The batches were also contaminated with exosomes containing HERV transcript RNAs (Zeilfelder *et al.*, 2006).

There are indications that HERVs could pose a threat when making LV or RV vectors for gene therapy. The potential side-effects of these HERVs on patients should be considered and investigated and with methodologies should to developed to filter these elements from gene therapy vector batches in order to produce safer gene therapy.

Chapter 2 - Materials and Methods

2.1 Materials

2.1.1 Antibodies and dyes

Alexa Fluor 488 goat anti-rabbit secondary antibody	Invitrogen
Alexa Fluor 488 rabbit anti-mouse tertiary antibody	Invitrogen
Anti-phospho-Histone H2A.X (Ser139), clone JBW301	Millipore
DRAQ5™ Fluorescent Probe Solution	Biostatus

2.1.2 Chemicals

6-Thioguanine, 98%	Thermo Fisher Scientific
8-Azaguanine 98%	Thermo Fisher Scientific
Acumax	Millipore
Blasticidin S	Thermo Fisher Scientific
Bovine Serum Albumin (BSA) in PBS (10%)	Thermo Fisher Scientific
DAPI (4′, 6-diamidino-2-phenylindole)	Thermo Fisher Scientific
DEAE Dextran	Thermo Fisher Scientific
Dimethyl Sulphoxide (DMSO)	Sigma Aldrich
Dulbecco's Modified Eagle Medium (DMEM)	Life Technologies
Ethanol	Thermo Fisher Scientific
Foetal bovine serum (FBS)	Life Technologies
HAT Supplement	Thermo Fisher Scientific
HT Supplement	Thermo Fisher Scientific
Nail Varnish	Essence Cosmetics
Natrium Chloride (NaCl)	Thermo Fisher Scientific
Paraformaldehyde Powder	Sigma Aldrich
PenStrep	Life Technologies

Phleomycin	Thermo Fisher Scientific
Phosphate buffered saline (PBS) in liquid	Life Technologies
Phosphate buffered saline (PBS) tablets	Thermo Fisher Scientific
Rabbit Serum	Millipore
SYBR® safe DNA gel stain	Thermo Fisher Scientific
TriTrack DNA loading dye	Thermo Fisher Scientific
Tris-Borate-EDTA (TBE) base	Thermo Fisher Scientific
Triton X-100	Sigma-Aldrich
Trypan-Blue	Thermo Fisher Scientific
Trypsin	Life Technologies

2.1.3 Commercial Kits

DNeasy blood & tissue	Qiagen
PureLink™ Quick Gel Extraction and PCR Purification Kit	Thermo Fisher Scientific
Long Range PCR Kit	Qiagen

2.1.4 Consumables

1.5 ml reaction tube	Thermo Fisher Scientific
10 cm dish	Thermo Fisher Scientific
10 ml pipette, sterile	Thermo Fisher Scientific
1000 ml glass bottles	Thermo Fisher Scientific
15 ml centrifugation tube	Thermo Fisher Scientific
175 cm^2 Tissue Culture flasks with filter cap	Thermo Fisher Scientific
25 ml pipette, sterile	Thermo Fisher Scientific
45 µm syringe filter	Thermo Fisher Scientific
5 ml pipette, sterile	Thermo Fisher Scientific
50 ml centrifugation tube	Thermo Fisher Scientific
6 well plate	Thermo Fisher Scientific

75 cm^2 Tissue Culture flasks with filter cap	Thermo Fisher Scientific.
Countess Cell Counting Chamber Slides	Thermo Fisher Scientific
Cryo tubes	Thermo Fisher Scientific
Eppendorf tube	Thermo Fisher Scientific
Pasteur pipettes	Thermo Fisher Scientific
Pipette tip with filter, 10 µl, sterile	Thermo Fisher Scientific
Pipette tip with filter, 1000 µl, sterile	Thermo Fisher Scientific
Pipette tip with filter, 200 µl, sterile	Thermo Fisher Scientific
Poly-L-Lysine coated PolyPrep Slides	Sigma-Aldrich
Polypropylene Microcentrifuge Tubes	Sigma-Aldrich
Quick-Load® 1 kb DNA ladder	New England Biolabs
Quick-Load® 1 kb Plus DNA ladder	New England Biolabs
Syringe Filters (0.45 µm)	Millipore
1.8 ml Cryotubes	Thermo Fisher Scientific

2.1.5 Equipment and Instruments

Axioscope 2 Imaging Microscope	Zeiss
Countess™ automated cell counter	Thermo Fisher Scientific
DM400 microscope	LCB
EVOS FLoid Imaging System	Thermo Fisher Scientific
Heraeus Biofuge Primo Centrifuge	Thermo Fisher Scientific
Image Stream^X	Amnis Corporation
Incusafe MCO-17AIC Incubator	Sanyo Biomedical
Micro Centaur centrifuge	MSE
Microscope CK2	Olympus
Molecular imager gel doc XRS	Bio-Rad Laboratories
Mr. Frosty™ Freezing Container	Thermo Fisher Scientific

NanoDrop2000 UV-Vis Spectrophotometer	Thermo Fisher Scientific
NovoCyte Flow Cytometer	ACEA Biosciences
Pasteur Pipettes	Thermo Fisher Scientific
Pipette gun, Accurpette	VWR
Pipettes, P10, P20, P200, P1000	Gilson
Rotator	Stuart
Sterile bench: Laminar Flow	Hera safe Thermo
Tabletop centrifuge (1-14)	Sigma Aldrich
Vibrax VXR Orbital Shaker	IKA
Vortex Mixer	Thermo Fisher Scientific
Water bath GD 100	Grant Instruments

2.1.6 Software

IDEAS®	Amnis Corporation
Image J	NIH (USA)
Image Lab v5.2.1	Bio-Rad Laboratories
Isis Fluorescence Imaging Software	MetaSystems
Metafer4	Zeiss
Microsoft Office 2010	Microsoft Corporation
Microsoft Paint	Microsoft Corporation
Multiple Primer Analyser	Thermo Fisher Scientific
NovoExpress software v1.2.5	ACEA Biosciences
Primer 3	ELIXIR
SnapGene 6.0	GSL Biotech LLC

2.1.7 Restriction Enzymes

EcoR32I/EcoRV	New England Biolabs
KpnI	New England Biolabs
Cut Smart Buffer	New England Biolabs

2.1.8 Bn Primers for the detection of HERV DNA sequences

All primers were obtained from Thermo Fisher Scientific

> **Table 2.1 Primers Used in Study.** Five primer pairs developed for a near 7.5 kb sequence within the regions not including the LTRs and each pair is labeled from H1-H5. 3 primer pairs developed for larger sequences which included the LTRs are labelled LTR1-LTR3

HERV Pair Name		Primer Name	Sequence (5' to 3')	Tm (°C)
H1		HERV Forward 1	CAACCCACCCCTACATCTGG	66.7
		HERV Reverse 1	CGACAAAACCACCATCGTC	64.5
H2		HERV Forward 2	TCTCTAGGGTGAAGGTACGC	61.3
		HERV Reverse 2	TGCTTTTCCCCACATTTCCC	68.2
H3		HERV Forward 3	CTCGAGCGTGGTCATTGAG	65.2
		HERV Reverse 3	CTCTCTTGCTTTTCCCCACATT	65.6
H4		HERV Forward 4	GGCTTTTCTCTAGGGTGAAGG	63.3
		HERV Reverse 4	GACAAAACCACCATCGTCATC	64.6
H5		HERV Forward 5	ACAAGTCGACGAGAGATCCC	63.7
		HERV Reverse 5	CGACAAAACCACCATCGTC	64.5
LTR1		HERV LTR Forward 1	ATTAAGGGCGGTGCAGGATG	68.3
		HERV LTR Reverse 1	TGATCATCTGTGGGTGTTTCTC	64.3
LTR2		HERV LTR Forward 2	GATCAGATTGTTACTGTGTC	52.2
		HERV LTR Reverse 2	AACCAGCGTTCAGCATATGG	65.3
LTR3		HERV LTR Forward 3	GAAGGCAGCATGCTCCTTAAG	65.3
		HERV LTR Reverse 3	GAGGATCCCGCCAGCCTCTGAG	74.5

2.1.9 Buffers and Solutions

2.1.9.1 ImageStream

Block Buffer: 5% Rabbit serum, 0.1% TritonX-100 in PBS

Permeabilisation buffer: PBS, 0.5% Triton-X 100

Fixing solution: 50% methanol, 50% Acetone

Washing buffer: 0.1% Triton-X 100 in PBS

2.1.9.2 ImmunoCytochemistry

Block Buffer: BSA 1% in PBS

Permeabilisation buffer: PBS, 0.5% Triton-X 100

TBST 137 mM NaCL, KCl 2.7 mM, Tris Base 19 mM

Fixing solution: 0.8% Paraformaldehyde in PBS

2.1.9.3 General

DMEM-Complete medium: DMEM + 2 mM L-Glutamine, 10% FBS, 1% PenStrep.

HAT medium: DMEM + 2 mM L-Glutamine, 10% FBS, 1% PenStrep, sodium hypoxanthine 10 mM, aminopterin 40 µM and thymidine 1.6 mM.

HT medium: DMEM + 2 mM L-Glutamine, 10% FBS, 1% PenStrep, sodium hypoxanthine 10 mM, thymidine 1.6 mM

1x Phosphate buffer saline PBS: For 1L: 8.0 g NaCl, 1.15 g Na_2HPO_4, 0.2 g KH_2PO_4, 0.2 g KCl

Paraformaldehyde Solution: 4% (w/v) paraformaldehyde in PBS.

2.1.10 Cell Lines

The cell lines used in this thesis were provided by Professor Michael Themis, Brunel University London, 2016 and Professor Christopher Parris, ARU, 2016.

MRC5SV1
SV40 Immortalised lung fibroblast capable of normal DNA following damage (Arlett *et al.*, 1988).

XP14BRneo17
SV40 Immortalised lung fibroblasts which are NHEJ defective and deficient in DNA PKcs (Abbaszadeh *et al.*, 2010).

FLYA-13
Epithelial like human fibrosarcoma produced for the production of replication incompetent murine leukaemia virus (MLV) for the purpose of gene therapy (Miller and Metzger, 2011). An altered form of the cell line lacking the MLV genome in the virus produced by the cell line was utilised.

PA317
Derived from NIH 3T3 TK- cells by co-transfection of the defective viral DNA. DNA construct consist of the promoter, *gag, pol* and *env* sequences of a helper virus useful for making retrovirus packaging cell line that do not transfer the packaging function

(Miller and Rosman, 1989). The virus produced by the cells was generated in mouse cells.

TELCeB/ AF-7
TELCeB/ AF-7 cells contain pCeB (*gag/pol*), pAF7 (amphotropic envelope) and the pMfGns LacZ backbone (Cosset *et al.*, 1995). These cells are referred to as 'AM7' in the study.

AT5BIVA
An immortalised fibroblast cell line derived from AT2SF with permanent ataxia-telangiectasia (Murnane, Fuller and Painter, 2003).

V79
Immortalised male neoplastic cell line from Chinese Hamsters (Ford and Yerganian, 1958)

NIH3T3
Referred to as 3T3. Embryonic mouse fibroblasts developed from NIH Swiss mouse embryo cultures (Jainchill, Aaronson and Todaro, 1969)

HEK293T
Referred to as 293T cells. Derived from HEK293 cells, embryonic human kidney cells with parts of DNA from human adenovirus type 5. The cells have the SV40 T antigen which improves transfection (Yuan et al., 2018; Shaw et al., 2002; Rio, Clark and Tjian, 1985).

STAR RD Pro
Referred to as SRH cells. Derived from the female foetal STAR cell line, which is used for packaging and producing retroviruses. The cell line permanently expresses *GFP*-LV (Ikeda *et al.*, 2003). Cell line was provided by Dr. Yasuhiro Takeuchi, UCL.

2.1.11 Plasmids for Restriction Digestion Tests

P54 – Plasmid FLAG-p54. Contains one site for the restriction enzyme KpnI ((Rosonina *et al.*, 2005).

P68 – Plasmid pBAD/HisB-iRFP682. Contains one site for the restriction enzyme EcoR32I/EcoRV (Scherbakova and Verkhusha, 2013).

2.1.12 Plasmid Vectors

Vector plasmids were primarily used to construct the vectors being provided by Prof. Michael Themis and Dr. Saqlain Suleman.

RHR'SIN-cPPT-SEW transfer vector used to make *GFP*-LV/pHR-LV (Demaison et al., 2002).

pCMVR8.74 was used for vector construction and contained the *gag* and *pol* genes (Dull et al., 1998). This shuttle vector is grown in bacteria using the necessary components for replication and expression.

pMD2.G was used for vector construction and contained the *VSV-G* envelope gene (Dull et al., 1998). This shuttle vector is grown in bacteria using the necessary components for replication and expression.

Figure 2.1 Schematic Showing the pCMVR8.74 Plasmid. The schematic shows the plasmid with the *gag* and *pol* regions along with other regions. The AmpR promoter drives *AmpR* expression required for ampicillin resistance for bacterial selection. The SV40 promoter, the CMV enhancer, CMV promoter and small t intron induce *gag* and *pol* expression in mammalian cells (Prasad and Rao, 2005; Isomura and Stinski, 2003; Huand and Gorman, 1990). The SV40 nuclear localization signal (NLS) is required for tagging proteins for import into the host nucleus (Collas and Aleström, 1996). The cPPT/CTS complex improves transduction of viral RNA into non-dividing human cell nuclei (Zennou *et al.*, 2001).

The RRE site is a binding site for the rev protein which is involved in regulating viral protein expression (Lensik, Sampath and Ecker, 2002). The schematic was generated using the SnapGene viewer software and obtained from Addgene, 2022a.
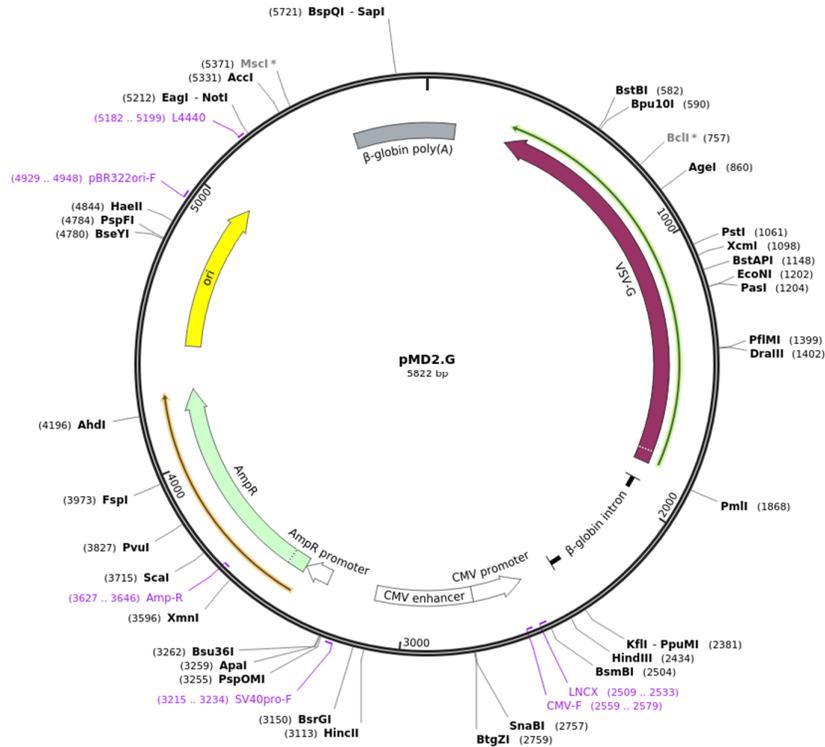
Figure 2.2 Schematic Showing the pMD2.G Plasmid. The schematic shows the plasmid with the *VSV-G* region (which codes for the viral envelope) and other regions. The AmpR promoter drives *AmpR* expression required for ampicillin resistance for bacterial selection. The CMV enhancer and CMV promoter induce *VSV-G* expression in mammalian cells (Isomura and Stinski, 2003). The β-globin intron contains a chimeric CMV enhancer and assists in enhancing gene expression (Haddad-Mashadrizeh, 2009; Nott, Le Hir, & Moore, 2004). The β-globin poly(A) signal allows for transcription termination (Yeung *et al.*, 1998). The schematic was generated using the SnapGene viewer software and obtained from Addgene, 2022b.

2.1.13 Viral Vectors

Viral Vectors were provided by Prof. Michael Themis and Dr. Saqlain Suleman.

GFP-LV: Known as pHR-LV. Provided after being made using the RHR'SIN-cPPT-SEW transfer vector. It contains the *GFP* gene with SIN LTRs (Demaison *et al.*, 2002).

Empty Vector: Was provided after being constructed using the pCMVR8.74 plasmid (**figure 2.1**), which codes for the *gag* and *pol* genes, and the pMD2.G plasmid (**figure 2.2**), which codes for the *VSV-G* envelope protein. The vector itself contains no genome.

2.2 Methods

2.2.1 Cell Culture

The cell lines utilised in this study (AM7, PA317, V79, AT5-BIVA, 3T3, 293T, XP14BRneo17 and MRC5SV1) were grown in specific media. Unless otherwise indicated, cells were grown in complete medium of Dulbecco's Modified Eagle Medium (DMEM) (Invitrogen Gibco) with 10% FBS (Invitrogen Gibco), 500 U/ml penicillin and 500 U/ml streptomycin (Invitrogen Gibco).

Cells were grown as monolayers in 92 mm plastic culture dishes (Thermo Scientific Nunc) in an incubator maintained at 37 °C with 5% CO_2 and a 95% air atmosphere humidified environment. Once cells had reached 80% confluence, the culture medium was aspirated with a glass Pasteur pipette (Fisherbrand) followed by washing with Phosphate Buffered Saline (PBS) (Invitrogen Gibco) at 37 °C to remove dead cell debris.

2.2.2 Trypsination

After medium was aspirated from cells growing on plates or slides, 1x trypsin in an EDTA solution (Invitrogen Gibco) was added in a volume of 1 ml to slides and a volume of 5 ml to plates following which they were incubated at 37 °C for 5 minutes to allow cells to detach. Trypsin activity was then stopped by the addition of 5ml of complete medium. This was followed by pipetting the cells up and down to generate a single cell suspension. This suspension was centrifuged at 1500 rpm for 5 minutes to pellet the cells and the supernatant was discarded by aspiration. Cells were then re-suspended in fresh complete medium. For passaging plates, a 1 in 5 dilution of the re-suspend cells were added to each new plate in 10 ml of total fresh medium. The cells were passaged 2 to 3 times a week depending on their doubling frequency or by observing when cells had reached 80% confluency. Cells grown on slides were trypsinised only for viability and were not passaged.

2.2.3 Cell Counting and viability check

Cells were counted using a glass haemocytometer or the Countess slides (Invitrogen). The haemocytometer was used following trypsinisation. A small sample of 10 µl was taken from a suspension of 10 ml and mixed with 10 µl of trypan blue following which cells were counted at 20x magnification in the Olympus CK2 microscope manually for cells at low concentrations. The Countess machine is able to count higher

concentrations (from around 1×10^4 - 1×10^7 cells/ml) with the protocol being the same of mixing 10 µl of the cell suspension with 10 µl of trypan blue. The machine automatically counts the number of live and dead cells and gives a ratio in the form of a percentage. For cells grown on slides, small samples of the cells seeded were grown alongside the slides and viablilities tested after the same testing agents were used on the sample cells at the same time points the cells grown on the slides were taken for ICC counts.

2.2.4 Cell Freezing

This was done to conserve stocks of cells. The freezing medium was composed of 10% FCS, 10% dimethyl sulphoxide (Sigma-Aldrich) and 80% DMEM medium containing penicillin/streptomycin and L-glutamine. Once cells were grown to around 95% confluency, they were trypsinised (as outlined above). Following recovery, they were centrifuged at 1500 rpm for 5 minutes and resuspended in freezing medium and placed in cryotubes. Cryotubes containing the cells were then soaked in isopropanol in a freezing container which facilitates a 1 °C/min gradual freezing to prevent cell damage. The cells were kept at -80 °C for 24 hours following which they were transferred to liquid nitrogen to allow for long term storage.

2.2.5 Slide Cell Growth for Immunocytochemistry

Upon reaching the correct confluency, the XP14BRneo17 and MRC5SV1 were trypsinised recovered in medium to a concentration of 200,000 cells/ml and 1 ml was spread across a poly-l-lysine coated slide. The slide was placed in a 92 ml petri dish and left overnight to grow. If infection was to be performed on a later date, the cells were topped up with complete medium to 10 ml.

2.2.6 Infection/Transduction

Cells to be infected were grown as noted in 92 mm tissue culture dishes or poly-l-lysine coated popyprep slides. Virus was either provided or obtained from cells producing viral vectors. Provided titrated virus ranged from concentrations of 1×10^{10} to 1×10^{12} vg/ml (viral genomes per ml) and were diluted to allow an MOI of 10, meaning a concentration of 10 virus particles for every cell to be infected was used. Virus obtained from producer cells was obtained by growing the producer cells to 95% confluency in an incubator at 37 °C and 5% CO_2 for 24 hours in fresh medium. The medium would then contain the virus produced at an unknown yet consistent concentration due to the standardised procedure of growing to a specific confluence and leaving the medium on for 24 hours.

For both titrated and producer obtained viruses, the virus was first complexed with 5µg/ml of DEAE Dextran (Thermo Fisher Scientific) for 20 minutes in 10 ml of complete medium in an incubator at 37 ^{o}C and 5% CO_2. This complexing assists with improving transduction. The growth medium on the cells to be infected was aspirated and replaced with the medium containing the virus. Plates were replaced in the incubator at 37 ^{o}C and 5% CO_2.

2.2.7 Fixation for ImageStream analysis

At specific time points, infected cells grown in plates were trypsinised and recovered with medium. These cells were then were washed with 5 ml of ice chilled PBS to kill the cells followed by centrifugation at 1500 rpm for 5 minutes. The PBS was aspirated and the cells were resuspended in 1 ml of methanol acetone in microcentrifige tubes and stored at -20°C until staining was to be done.

2.2.8 Fixation for Immunocytochemistry

The cells grown on slides were taken and the medium was aspirated and replaced with ice chilled PBS which was then aspirated. The slides were then placed in a plate and covered with 5 ml of 4% formaldehyde and 45 ml of PBS for 15 minutes. Following this the slides were removed and placed in 10 ml of PBS at 4°C until staining was to be done.

2.2.9 γH2AX staining for ImageStream analysis

Following the fixation describe in section 2.2.7. The cells were all handled in microcentrifige tubes. 0.25 ml of the primary antibody (Millipore) was added to the tube at a 1:10000 dilution in blocking buffer overnight at 4°C. The cells were then washed with 1ml of washing buffer three times on a rotator to allow for even washing of the cells. The secondary antibody was then added at a dilution of 1:1000 in blocking buffer. The cells were incubated for 1 hour at room temperature on a rotator with an aluminium foil covering it to ensure light does not damage the fluorescent material. The cells were washed 3 times once more in washing buffer. 20 µl of Accumax solution was then added to these tubes and left overnight at 4°C. DRAQ5 was then added to compensation tubes for each cell line to allow for the image stream analysis.

2.2.10 Immunocytochemistry staining to detect γH2AX and DNA damage

The slides were washed 3 times for 5 minutes each in TBST and agitated on an orbital shaker (IKA). They were then washed in 0.2% Triton at 4°C for 5 minutes to allow for

permeabilisation. The primary antibody (Millipore) was diluted in blocking buffer at 1:1000 and added to the slides at 100 µl per slide. The slides were covered in parafilm and left for an hour at room temperature. The slides were washed 3 times in TBST on the shaker. The secondary antibody was added at a 1:1000 dilution in blocking buffer at 100 µl per slide to each slide then covered with parafilm and left at room temperature for 30 minutes. The cells were rinsed in PBS on the orbital shaker once and covered with the tertiary antibody at a 1:1000 dilutions in blocking buffer for another 30 minutes. The slides were washed in TBST 3 times then in PBS 3 times as mentioned on the shaker. The slides were dehydrated sequentially by adding increasing concentrations of ethanol at 70%, 90% and 100% for 5 minutes each. The slides were air dried and 15 µl of DAPI was added to the slides. A coverslip was placed on the slide ensuring no bubbles remained following which clear nail varnish was used to seal the cells.

2.2.11 ImageStream Analysis for DSB

The template for the analysis was loaded and the folder to which analysis files would be saved was selected. The machine was then prompted to take up the sample. The samples only containing one specific dye (Alexa Fluor or DRAQ5) with each cell line were first run through the machine to allow the machine to establish a compensation matrix to reduce background for the ImageStream analysis. The actual samples were then run through the machine which counted the foci indicating γH2AX localisation, (thus indicating DNA damage) and provided a cumulative mean for 10,000 cells. Data for each sample was obtained and saved containing information for at least 10,000 cells when possible.

Analysis of the data was performed in the IDEAS software. First, the compensation matrix was loaded to determine the amount of background fluorescence. The Spot wizard was then loaded and the channels containing the DRAQ5, Alexa Fluor and Images of the cells were selected for analysis out of the five possible channels which the machine allows for analysis (the other two were not necessary since only two dyes were used). To ensure the machine only counted single cells in focus, parameters were established to exclude clumped up cells mistakenly counted as single cells and cells out of focus. The channel with the green Alexa Fluor dye indicating the DNA damage was selected to be analysed and 'truth' populations were established to allow the machine to differentiate cells with few foci and those with many by indicating cells with only one focus and those with many distinct foci. The machine was then prompted to analyse the data, while excluding samples based on the parameters provided, and cumulative mean

showing DNA double strand breaks across all cells was provided. This value was noted and used to determine the results.

2.2.12 Immunocytochemistry Analysis

The ISIS fluorescence imaging software in conjunction with the Axioskop 2 microscope was used to analyse the slides. The slides were observed at 10X to determine areas of high confluence on the slides then focused with oil immersion at 100X where a drop of oil was placed on the slide to prevent the refraction of light through air preventing the observation of cells at high magnification. 100 cells were observed and the foci of high intensity of green Alexa Fluor were counted to determine the points of double strand breaks. A cumulative mean was obtained by dividing the number of foci across all cells by 100. This value was noted and used to determine the results.

2.2.13 HAT Selection – HPRT Assay

HAT supplement (Thero Fisher Scientific), containing **h**ypoxanthine, **a**minopterin and **t**hymidine, were added to DMEM to produce HAT medium with concentrations of sodium hypoxanthine at 10 mM, aminopterin at 40 µM and thymidine at 1.6 mM.

HAT selection was then performed on cells by growing them in HAT medium at 37 ^{o}C and 5% CO_2 in an incubator. The cells were regularly passaged in HAT medium and counted until they numbered at least 1 x 10^7.

These cells could then be transfected with virus and undergo a secondary selection in 8AG/6TG.

2.2.14 8AG/6TG Selection – HPRT Assay

Cells were grown in in accordance with the cell culture protocol with an addition of 8AG to the solution to form a 30 µg/ml concentration. 6TG was also used interchangeably with 8AG at a concentration of 10 µg/ml. Cells were grown over 1-2 weeks to the point of clear and countable colony formation. Aspiration and replacement of medium with either 8AG or 6TG was done every 24 hours.

2.2.15 DNA Extraction and Purification

DNA was extracted and purified from cells according to the manufacturer's instructions using the DNeasy kit (Qiagen). Up to 5 x 10^6 cells were grown for extraction. Cells were suspended with proteinase K in 200 µl of PBS at a 1:10 ratio and incubated at 56°C for 10 minutes. The DNA was precipitated using ethanol and loaded into the

columns provided. The DNA was washed and eluted into the buffer provided. The purity and concentration of DNA was determined using the NanoDrop2000 UV-Vis Spectrophotometer and stored at -80°C until required.

2.2.16 Gel Electrophoresis

Agarose gels were prepared at 0.6%, 1% and 2% in 100ml 1X TBE for five minutes. The solution was cooled and, prior to setting, 1 µl 10,000X SYBR safe DNA gel stain was added before casting the gel. The gels were then placed in an electrophoresis tank and covered in 1x TBE. TriTrack dye was added to the nucleic acid samples before loading in wells alongside DNA ladders as required. Gels were run at 70V for 35-45 minutes to allow for band separation. Gels were imaged on the molecular imager gel doc XRS and analysed via the Image Lab software.

2.2.17 PCR to detect HERV Sequence

Primers were obtained from Thermo Fisher scientific in lyophilised form. They were hydrated in dH_2O to create 10mM stock solutions. Stock solutions were diluted to a 10µM working solutions for PCR reactions. Nucleic acid samples were added to a 50 µl or 100 µl for large scale reactions when indicated according to the manufacturer's instructions with the Long Range PCR Kit (Qiagen). Reactions included LongRange PCR buffer diluted to 1x from the 10x stock provided, Mg^{2+} at 2.5 mM), dNTP mix at 500 µM of each dNTP, LongRange PCR Enzyme mix at 2 units per 50 µl, forward and reverse primer at 0.4 µM each and RNase-free water to make the solution up to 50 µl or 100 µl. When Q-solution was utilised for optimisations, it was diluted to 1x from the provided 5x solution. Note that the exact contents and concentrations for the LongRange PCR buffer, PCR enzyme mix and Q solutions are proprietary information which is not shared with the user. The samples were amplified in a thermocycler using the program provided by the manufacturer with the following specifications. An initial denaturation of the template DNA was done at 93°C for 3 minutes. This was followed by a 3 step cycle repeated for 35 cycles. One cycle was as follows: i) denaturation at 93°C for 15 seconds, ii) annealing at 62°C for 30 second, and iii) extension at 68°C for 1 min/kb of sequence to be amplified. Extension ranged from 7 minutes and 30 seconds to 8 minutes and 30 seconds depending on the sequence. At the end of cycling, the product was held at 4°C until required. Annealing temperatures changed when mentioned.

2.2.18 Gel DNA Band Extraction and Purification

The manufacturer's instructions were followed using the PureLink™ Quick Gel Extraction and PCR Purification Combo Kit to extract and purify the DNA from the agarose. Following electrophoresis, a scalpel was used to carefully remove the desired band from the gel. The gel slice was then placed in a polypropylene microcentrifuge tube. Gel solubilisation buffer was added at a 6:1 ratio (for example 2.4 ml buffer for a 400 mg gel slice) and placed in a water bath at 50°C for 10-15 minutes with it being inverted every 3 minutes. Following visible gel dissolution, a further 5 minute incubation was done. Isopropanol was added in μl at 1:1 to the amount of gel slice being purified in mg, for example, 50 μl of isopropanol for 50 mg of the gel slice. The solution was then added to Clean-Up spin columns provided and centrifuged at 10,000 g for 1 minute. The flow-through was discarded and the column was placed in a wash tube and centrifuged at 10,000 g for a further 3 minutes to complete discard residual wash buffer and ethanol. The wash tube was discarded and the column was then placed in an elution tube. 50 μl of elution buffer (E1), or TE buffer was added to the column and incubated for 1 minute at room temperature before being centrifuged at 10,000 g for 1 minute. The tube retrieved would contain the amplified PCR product which was stored until required. The column was discarded.

2.2.19 Restriction Digest

Restriction enzyme digestion was used on the HERV PCR products to verify the sequence of the DNA in question. 500 ng of product following gel band extraction and purification was incubated with 1 μl of CutSmart buffer and 1 μl (1 unit) of each restriction enzyme KpnI and Eco32I (EcoRV) together or each in isolation made up to 10 μl with dH_2O. The digestion was done at 37°C for 1 hour. 1 unit of KpnI is described as the amount required to digest 1 μg of pXba (adenovirus fragment in plasmid) in 1 hour at 37°C in a total volume of 50 μl. 1 μg. 1 unit of EcoRV is the amount required to digest 1 μg of γ DNA (from bacteriophage lambda) in 1 hour at 37°C in a total volume of 50 μl.

2.2.20 Statistical Analysis

Comparisons of two sample groups were analysed using the student's T-test to generate P values to determine the significance of the differences. Statistical significance was done at 95% and 99% confidence intervals. Standard error of the mean (SEM) was also used. Any graphical representation was done using Microsoft Excel 2013.

Aims and Objectives - Overview

Retrovirus (RV) genotoxicity firstly involves integration of the vector into the host DNA. NHEJ is required for the integration process to repair DSBs. This thesis aims to test the hypothesis of three main aspects of viral integration in host cells:

1. That infection would likely cause DNA damage and genotoxicity at higher levels in cells that do not have an intact NHEJ pathway compared to cells that do.

2. That infection can cause damage even with "empty" vectors lacking genetic material to be integrated.

3. That if empty vectors are causing damage, it may be due to HERVs being incorporated into them.

Chapter 3 – γH2AX DNA Damage Profile Comparison on NHEJ Competent and NHEJ Compromised Mammalian Cells

3.1 Background and Aims

For retroviral gene therapy vectors to insert genetic sequences into target cells, the vectors make DSBs in the target cells' DNA using the viral protein, integrase. This allows the vector to use this DSB to allow for the integration of the therapeutic DNA sequence. During this integration, the target cell's own NHEJ mechanism is utilised to reseal the DSBs along with the therapeutic sequence.

3.1.1 Aims and Objectives

- To determine the DNA damage repair profile of cells with or without the NHEJ pathway using assays that detect the emergence and disappearance of phosphorylated histone γH2AX, an indicator of DSBs.

- To use retroviruses either with or without vector genomes to determine whether the virus genome is necessary for DSBs to occur

- To determine the level of residual DSBs by measuring the persistence of γH2AX following infection in cells with or without the NHEJ pathway

To test the hypothesis that the NHEJ pathway is necessary to extensive damage in the host cell DNA, cells with and without functioning NHEJ mechanisms were infected and their DNA damage profiles prior to and following infections were determined.

To determine whether the vectors need to carry a vector genome to cause DSB damage when infecting cells, the viruses used in the prior paragraph were LV vectors containing a therapeutic genome and "empty" vectors lacking a genome.

To determine the extent of DSB formation and repair, a γH2AX assay was employed. H2AX is a histone protein which becomes phosphorylated once a DSB occurs and is referred to as γH2AX. Primary and secondary antibodies sensitive to γH2AX can be paired with a fluorescent marker that can be detected via cell cytometry or immunohistochemistry as seen in **figure 3.1**. This allows for the detection of individual instances of double strand breaks forming in cells, allowing for a quantifiable assay.
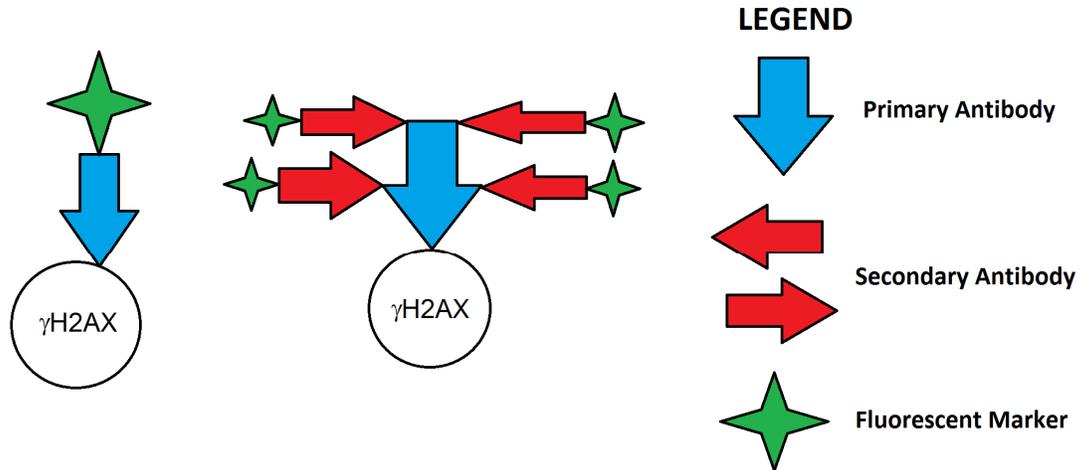
Figure 3.1 γH2AX Secondary Antibody Staining Principle. While primary antibodies tagged with a marker can be used to bind to a protein of interest, such as γH2AX, only a single marker will denote the position of the protein; this may produce a faint signal which may be difficult to detect. Whereas when secondary antibodies tagged with a marker which can attach to a primary antibody are used, for each primary antibody which attaches to the protein of interest, there will be several fluorescent markers which will produce a stronger signal for detection. Image constructed in Microsoft Paint.

3.2 Previous supporting data to the work

3.2.1 Prior Immunocytochemistry Data

Previous work carried out by a PhD student, Safia Reja was carried out in our laboratory. This involved immunocytochemistry to compare the γH2AX DSB profiles of the NHEJ impaired and NHEJ non-impaired cell lines, MRC5-SV1 and the XP14BRneo17 respectively. The results of this work can be seen under **appendix 1**.

The results showed a more profound DNA damage profile for cells infected with the MLV compared to the level of damage seen following irradiation demonstrating the potential hazard of using gene therapy for patients who might have defective DNA repair. Viruses lacking the genome showed increased initial damage compared to MLV carrying a genome.

This study focuses at obtaining more data in order to observe the effects of infection of cells with viruses which can be utilised as vectors over more time points.

3.2.2 Prior ImageStream Data

Other prior work was conducted over more time points following MLV and HIV-F9 infection, both of which are utilised for gene therapy, on the NHEJ competent and NHEJ incompetent cell lines, MRC5SV1 and XP14BRneo17. ImageStream, an advanced flow cytometer with an imaging device which can observe a high throughput of individual cells in high resolution, was used to process the cells and count the DSBs via γH2AX DSB profiles. The results of these experiments are seen in **appendix 2**. The data showed that the NHEJ-impaired cell line sustained a higher amount of DSBs.

This existing data shows a potential hazard to patients who may have defective DNA damage repair mechanisms if treated with gene therapy vectors. Thus, it is necessary to determine the possible damage this can do and as such more data were gathered for the study.

3.3 ImageStream of γH2AX to Determine DNA Damage Repair

ImageStream was first utilised in order to obtain large amounts of data with each run as described in the methods section. Five tissue culture dishes of the MRC5 and XP14 cell lines were grown to 95% confluency and infected with MLV virus lacking a genome and the HIVF9 virus as describes in the materials and methods **section 2.2.5**. The infected cells were then harvested, fixed and stained for γH2AX phosphorylation followed by ImageStream processing as described in **section 2.2.6** and **2.2.8** at the time points stated. The control time point is the number of γH2AX at 0 hours, or before infection.

3.3.1 Infection of Cells by Human Produced MLV with Genome
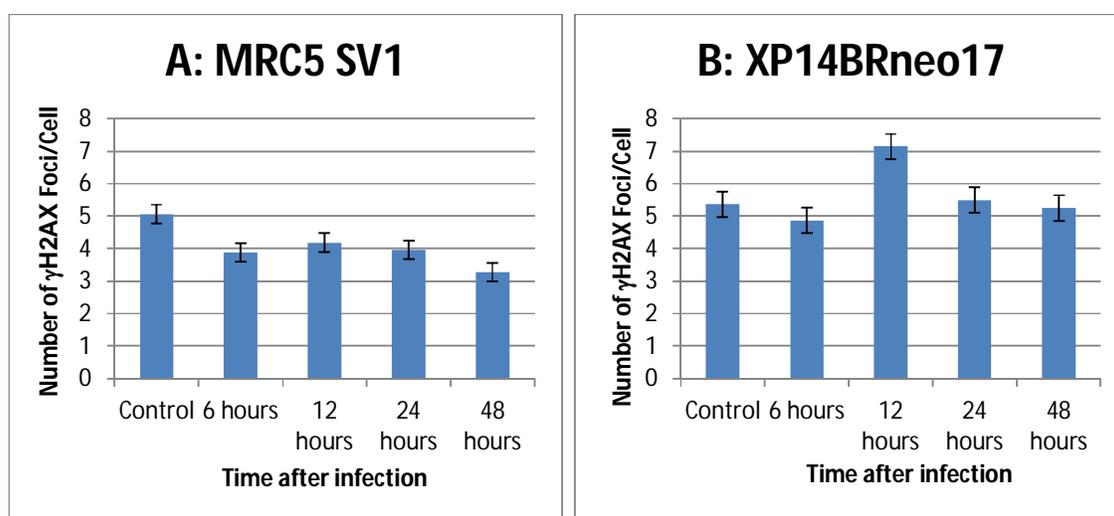
Figure 3.2 MRC5-SV1 and XP14BRneo17 γH2AX DNA damage profiles post-MLV transduction.
The number of γH2AX foci per cell indicate the mean of DSBs in a collective amount of cells. **A**) shows the MRC cell line profile while **B**) shows the XP14 cell line profile. Both cell lines show an initial high level of γH2AX at the control, or the 0 hour time point compared to the 6 hour time point following which the foci fluctuated. MLV was obtained from the AM7 producer cell line. Infection was done with a consistent titre on both cell lines as described in the methods **section 2.2.6**. Error bars represent ±SEM.

The MRC profile (**figure 3.2 A**) and XP14 profile (**figure 3.2 B**) show a high number of foci for the control at the 0 hour control time point as compared to the 6 hour time point. This is incongruent to the previous data from prior experiments as shown in the appendix sections. This high control value was not expected, as prior to infection, the DNA should be the most stable and thus having a low number of DSBs. As expected, however, the MRC5 cell line with an intact NHEJ mechanism had fewer DSBs across all time points as compared to the DNA repair impaired XP14 cell line.

3.3.2 Infection of Cells by Human Produced MLV without Genome
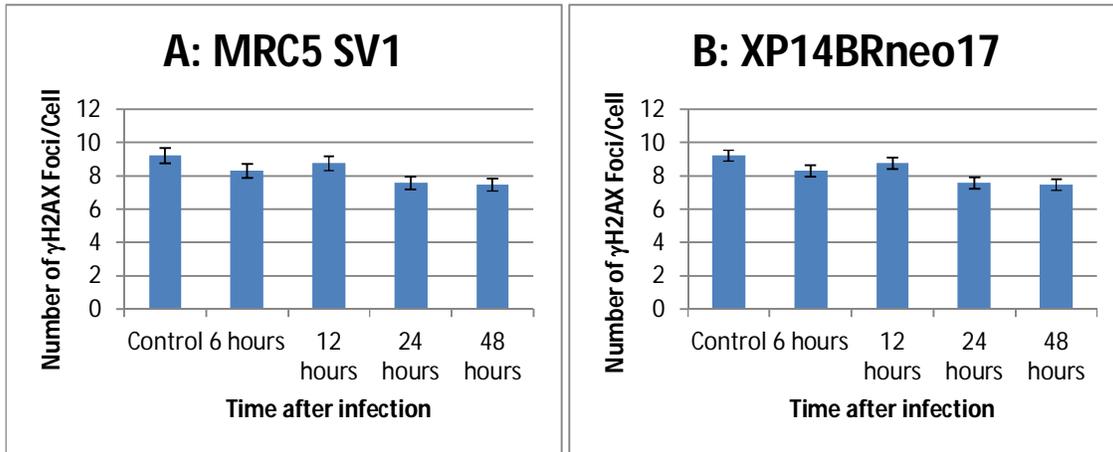
Figure 3.3: MRC5-SV1 and XP14BRneo17 γH2AX DNA damage profiles post-empty human-MLV transduction. The number of γH2AX foci per cell indicate the mean of DSBs in a collective amount of cells. **A**) shows the MRC cell line profile while **B**) shows the XP14 cell line profile. Both cell lines show an initial high level of γH2AX at the control, compared to the 6 hour time point following which the foci fluctuated. MLV was obtained from the FLYA13 cell line. Infection was done with a consistent titre on both cell lines as described in the methods **section 2.2.6**. Error bars represent ±SEM.

The MRC profile (**figure 3.3 A**) and XP14 profile (**figure 3.3 B**) show a high number of DSB foci at the 0 hour, control mark similar to the results seen in **section 3.3.1**. This time, unlike in **section 3.3.1**, the number of foci are similar across both cell lines. This incongruence with the prior work done as well as the prior experiment in **section 3.3.1** shows an error in either the infection methodology, an error in the ImageStream processing or an error in the ImageStream analysis.

3.3.3 Infection with Mouse Produced MLV without Genome
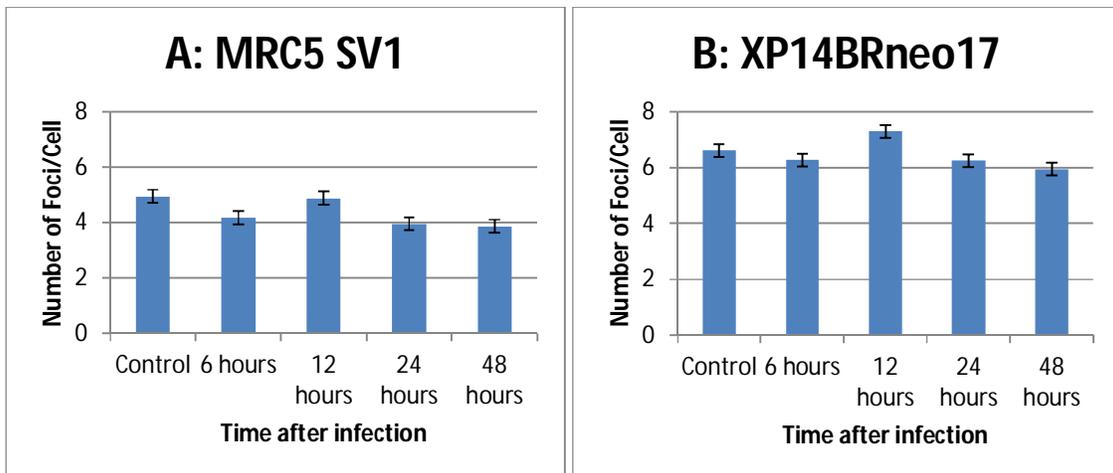
Figure 3.4 MRC5-SV1 and XP14BRneo17 γH2AX DNA damage profiles post-empty mouse-MLV transduction. The number of γH2AX foci per cell indicate the mean of DSBs in a collective amount of cells. **A**) shows the MRC cell line profile while **B**) shows the XP14 cell line profile. Both cell lines show an initial high level of γH2AX at the control, or the 0 hour time point compared to the 6 hour time point following which the foci fluctuated. Across all time points, the MRC cell lines showed a lower amount of DSBs than the XP14 cell line. MLV was obtained from the PA317 cell line. Infection was done with a consistent titre on both cell lines as described in the methods **section 2.2.6**. Error bars represent ±SEM.

The MRC profile (**figure 3.4 A**) and XP14 profile (**figure 3.4 B**) show a high number of DSB foci at the 0 hour, control mark similar to the results seen in **section 3.3.1 and 3.3.2**. Like in **section 3.3.1**, the MRC5 cell line with an intact NHEJ mechanism had fewer DSBs across all time points as compared to the DNA repair impaired XP14 cell line. The consistent high 0 hour control values compared to the 6 hour makes indicate a persisting error across experiments.

3.3.4 Infection with HIVF9 with Genome
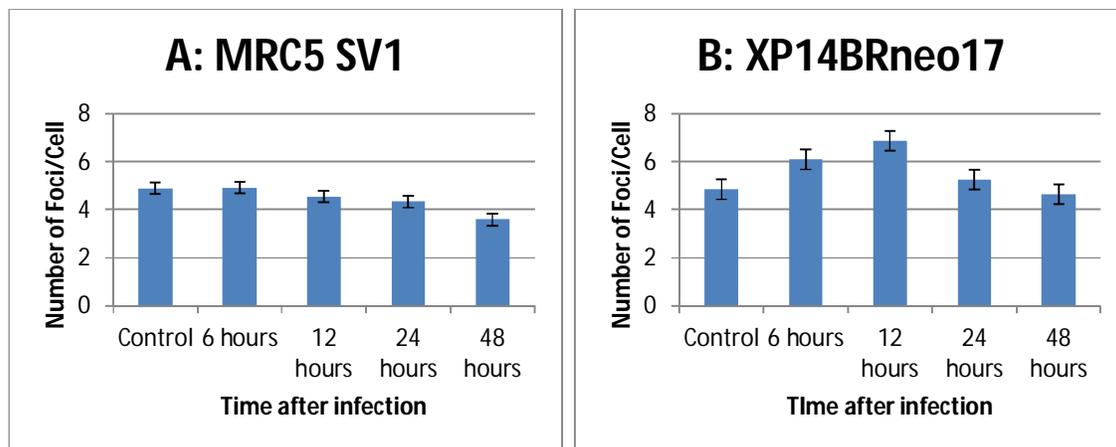
Figure 3.5 MRC5-SV1 and XP14BRneo17 γH2AX DNA damage profiles post-MLV transduction. The number of γH2AX foci per cell indicate the mean of DSBs in a collective amount of cells. **A**) shows the MRC cell line profile while **B**) shows the XP14 cell line profile. Infections were done at an MOI of 10 with HIV-F9 viral aliquots provided by Prof. Mike Themis. Error bars represent ±SEM.

This experiment was done with viral aliquots of known titre at an MOI of 10 while prior experiments were done with consistent titres that were unknown. The MRC profile (**figure 3.5 A**) had a similar number of DSBs at the 0 hour control time point as compared to the 6 hour time point followed by a gradual decrease in DSBs until the 48 hour time point. This is unexpected as prior work indicates there should be a sharp increase in DSBs at the 6 hour mark followed by gradual repair. The lack of a sharp increase after infection shows incongruence with previous work. The XP14 profile shows

a more expected result with an increase in DSBs, peaking at the 12 hour mark, followed by a gradual decrease.

3.3.5 Reanalysis of ImageStream Raw Date from Prior Works

Most of the ImageStream experiments showed a high number of background DSBs in the control samples in both cell lines, MRC5 (which are NHEJ competent) and XP14 (which are NHEJ impaired) the exception being the XP14 profile (**figure 3.5 B**) seen in the infections done with the HIVF9 aliquots. Following this, certain runs showed the amount of DSBs decreasing after the control time point as seen in **figures 3.2 A & B, 3.3 A & B, and 3.4 A & B**. This indicates that the amount of DSB damage is seen to be decreasing following infection. As mentioned, these results are incongruent with previous work done as seen in the appendix section. These results go against the theory that infections cause DSBs which are then repaired. However, the results are not providing a consistent pattern as certain experiments showed results more congruent with prior work, as seen in **figure 3.5 B** where there is an increase in DSBs following infections, in line with our theory.

These inconsistent results indicate an error in either the experimental phase, or the data analysis phase. As such, to determine if there was an error in the analysis phase, the raw data from the prior ImageStream results as seen in **appendix 2** were re-analysed.
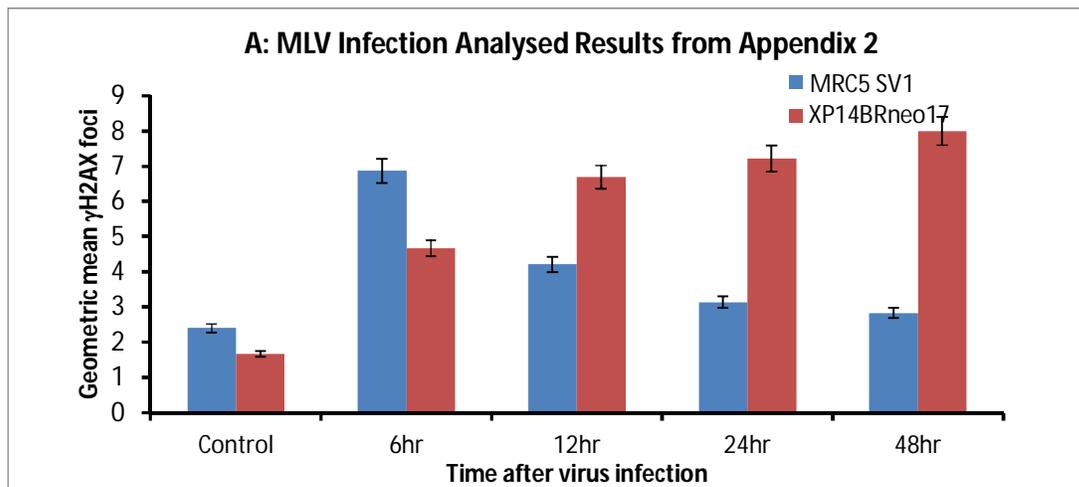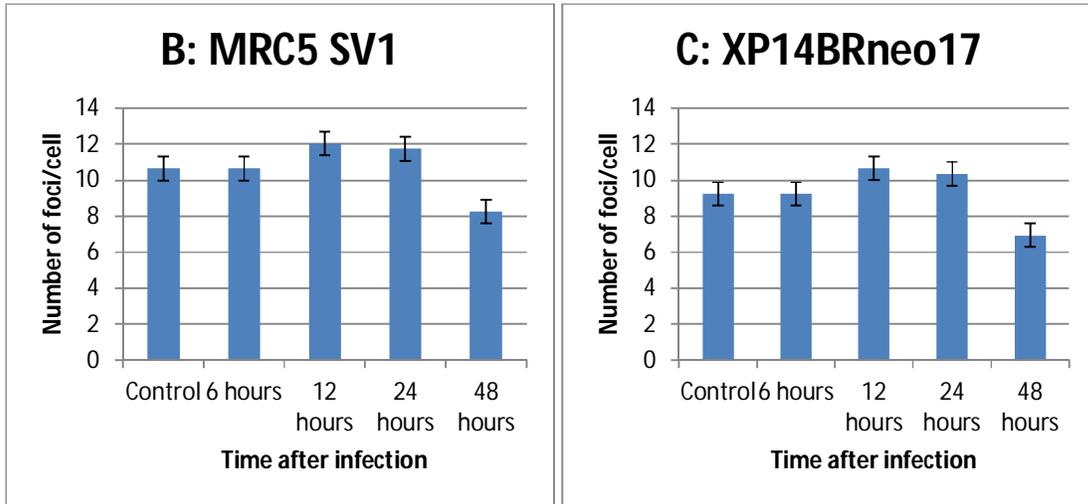
Figure 3.6: Reanalysis of MLV infection ImageStream raw data from Appendix 2. The number of γH2AX foci per cell indicate the mean of DSBs in a collective amount of cells. **A)** shows the graph of the results post analysis of raw data obtained from prior experiments as seen in **appendix 2**. **B)** and **C)** show the MRC5 line and XP14 line profiles respectively following analysis of the raw data used to generate graph **A)** from prior work done. A high number of DSBs is seen in the controls in **B** and **C** similar to what was seen in **figure 3.2-3.5**. Error bars represent ±SEM.

The results obtained in figures **3.6 B** and **C** are both incongruent from the results obtained in **figure 3.6 A** though both sets of results were generated following analysis of the same raw data. This indicates an issue with the raw data analysis step being a possible source of the incongruencies seen **figures 3.2 – 3.5**.
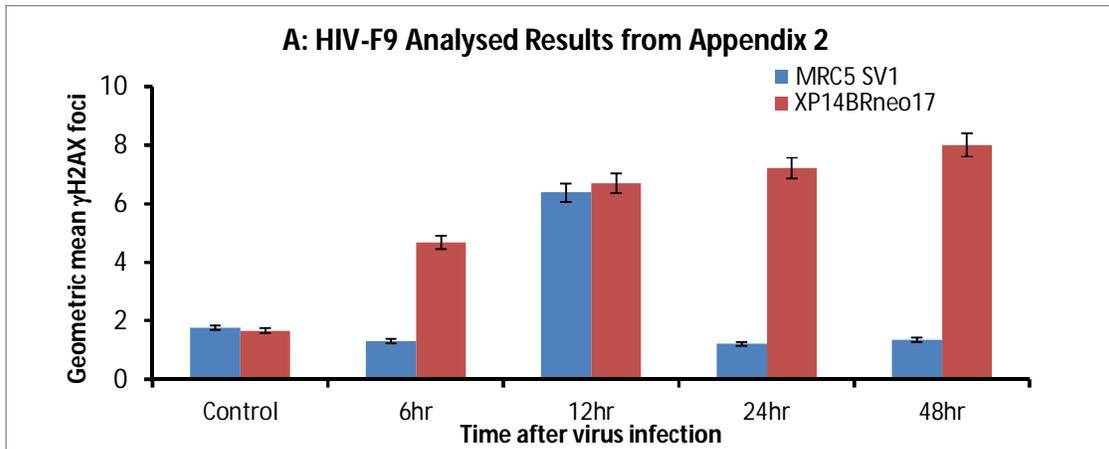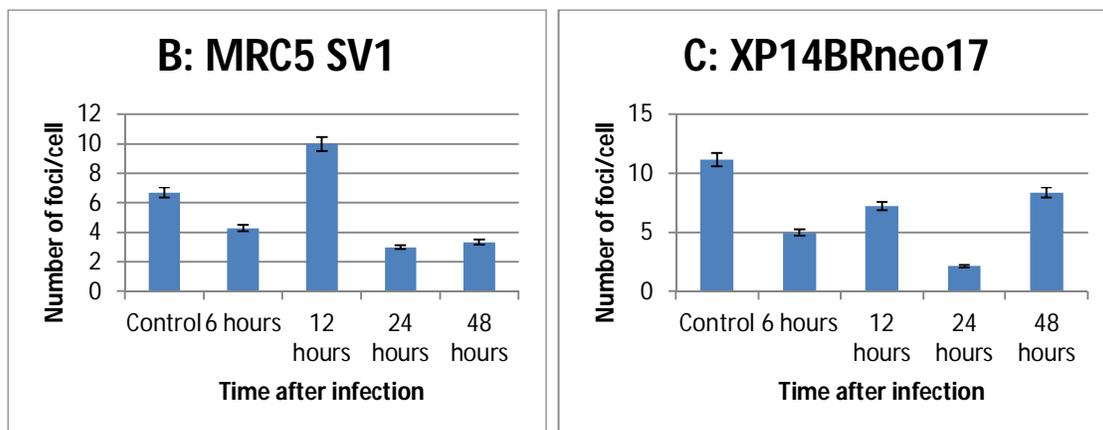
Figure 3.7: Reanalysis of HIVF9 infection ImageStream raw data from Appendix 2. The number of γH2AX foci per cell indicate the mean of DSBs in a collective amount of cells. **A**) shows the graph of the results post analysis of raw data obtained from prior experiments as seen in **appendix 2**. **B**) and **C**) show the MRC5 line and XP14 line profiles respectively following analysis of the raw data used to generate graph **A**) from prior work done. A high number of DSBs is seen in the controls in **B** and **C** similar to what was seen in **figure 3.2-3.5**. Error bars represent ±SEM.

As seen from the reanalysis in **figure 3.6**, figures **3.7 B** and **C** are both incongruent from the results obtained in **figure 3.7 A** though both sets of results were generated following analysis of the same raw data.

This indicates that there is at least an error in the ImageStream analysis done to obtain the results seen in figures **3.2-2.5**. It appears that the same error is occurring during analysis done for **figures 3.6 and 3.7** as these new results are indicating a high number of DSBs formed in the 0 hour control samples.

Due to the inability to reproduce the graphs shown in **appendix 2** following reanalysis of the raw data, it was determined that an alternative method to ImageStream should be attempted to be used for this study for γH2AX foci counts.

3.4 Immunocytochemistry of γH2AX to Determine DNA Damage Repair

The ImageStream data were found to have high background levels of DNA damage. As this made it difficult to determine the effects of virus infection on these cells an alternative method, immunocytochemistry, was used for this work.

The Immunocytochemistry method works by fixing cells on glass slides and using antibodies which attach to the desired γH2AX loci which can then be manually counted under a microscope. The foci appear as shown in **figure 3.8**.
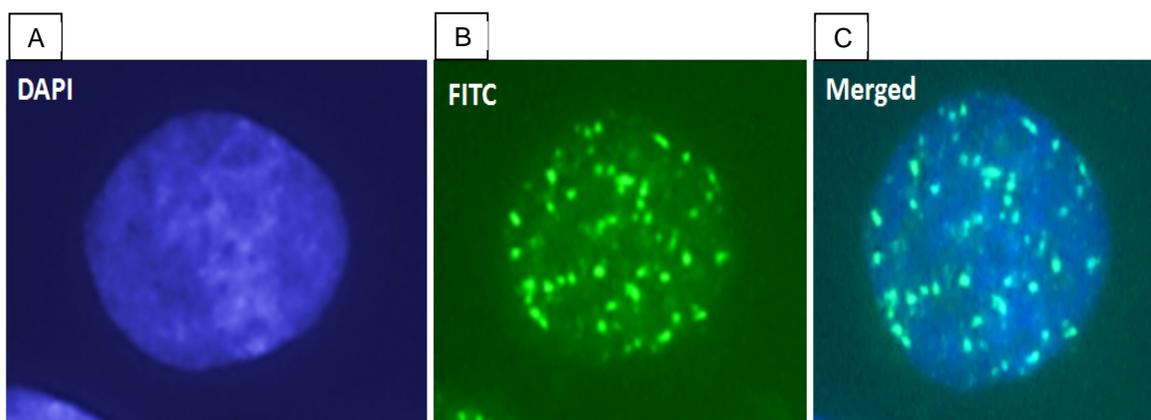
Figure 3.8: An example of γH2AX Damage profile of AT5BIVA Cells. Cells following immunocytochemical treatment to detect phosphorylated γH2AX loci, an indicator of DSB formation. **A)** The DAPI channel indicates DNA in blue. **B)** The FITC channel indicates green fluorescence which is emitted by the antibodies bound to the γH2AX loci, thus indicating DSBs. **C)** The merged channel overlays the locations of the DAPI and FITC staining to confirm the individual FITC loci fall within the area where DNA is present. The individual foci can be counted to quantify DSB formation. Images were viewed using the Zeiss Axioplan 2 Imaging Microscope and images were captured using the Metafer4 software.

Five polyprep slides of each of the cell lines with and without DNA DSB repair mechanisms, MRC5-SV1 and XP14BRneo17 respectively, were infected using MLV and HIV either with or without a genome as described in materials and methods **section 2.2.5**. The slides were then fixed and stained on slides by first exposing them to 0.2mM H_2O_2 and exposure to primary and secondary γH2AX antibodies in line with the methods outlined in the materials and methods **section 2.2.7** and **2.2.9**.

Cell viabilities were taken of cells in petri dishes which were infected at the same time as the cells on the slides at every time point as described in materials and methods **section**

2.2.3 to demonstrate that a majority of live cells were analysed, and thus a majority of DSBs formed were not due to cell death.

γH2AX were counted in a minimum of 100 cells for each sample. The cumulative mean of γH2AX foci was plotted against the time points the cells were fixed at.

3.4.1 AM7 Human Cell Line Produced Complete MLV Infections

The first set of infections were performed using genome carrying "complete" MLV (with viral genome) procured from the AM7 cell line as described in the materials and methods **section 2.2.5**. Consistent titres were used on each cell line.

3.4.1.1 Complete MLV Infection Set 1
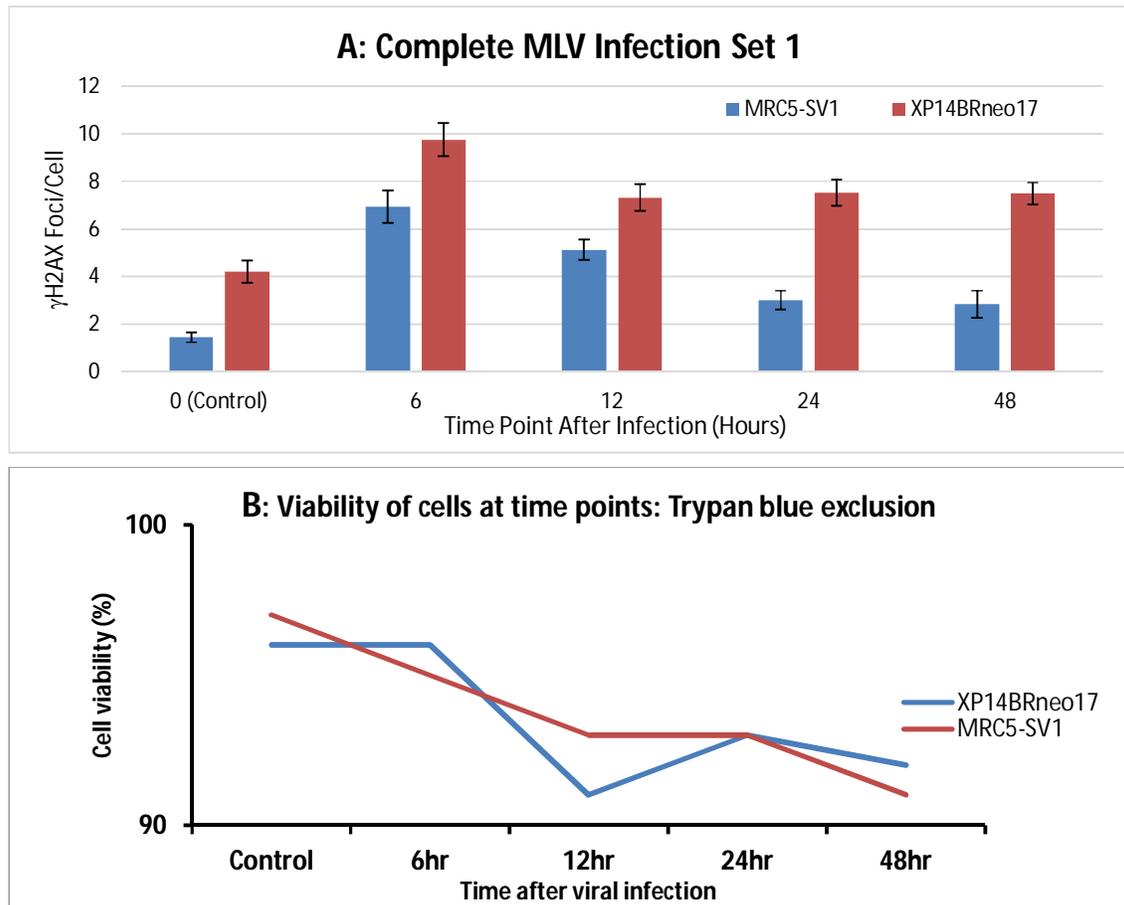
Figure 3.9 ICC γH2AX Damage Profile 1 of MRC5 and XP14 following MLV transduction. Immunocytochemistry analysis of 100 cells following infection and fixed at known time points. **A)** shows the cumulative mean of γH2AX foci per cell which indicate the mean of DSBs in a collective amount of cells of the MRC5 and XP14 cell lines. **B)** shows the cell viabilities of both cells lines staying above 90%

for all time points for both cell lines. Consistent titres were used on each cell line. Error bars represent ±SEM.

Compared to the ImageStream analysis results in **section 3.3**, the results seen with the ICC analysis on slides produce results which are in line with prior work done as seen in a**ppendix 1** and **appendix 2**. As seen in **figure 3.9 A**) Both cell lines have an increase in DSB formation at the 6 hour mark with a lower initial DSB count at the 0 hour control time point (unlike the ImageStream results which showed a higher DSB count at the 0 hour control time point). The MRC5, with intact DSB repair mechanism, shows DSB numbers falling up to the 24 hour mark and plateauing, while the XP14 line, with impaired DSB repair, shows an initial decrease up to the 12 hour time point at a much higher number of DSBs then plateaus, indicating less repair.

3.4.1.2 Complete MLV Infection Set 2
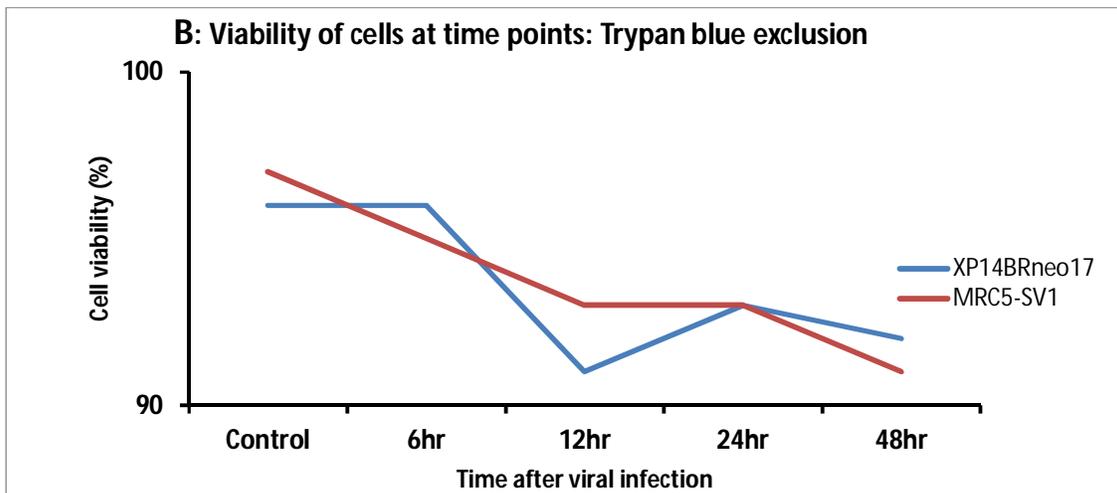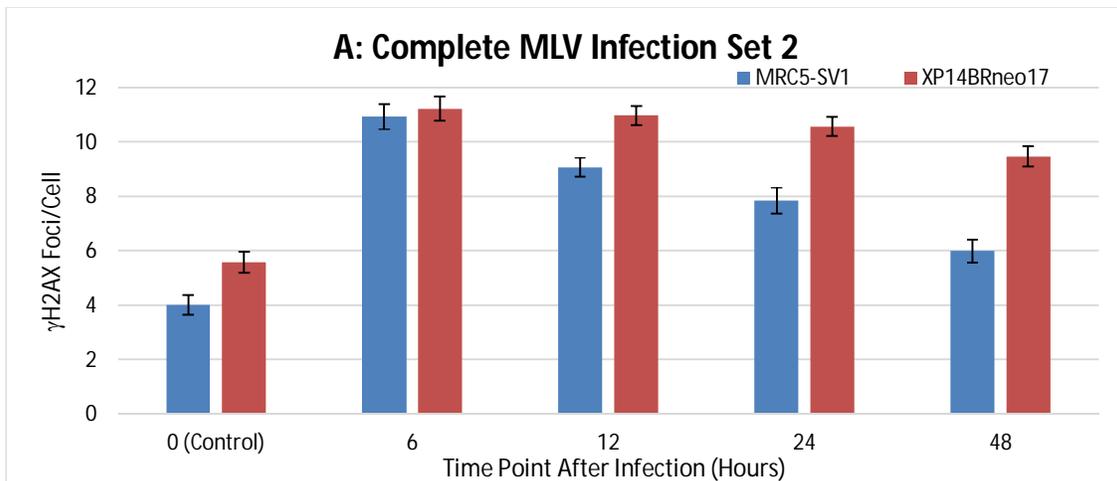
A: Complete MLV Infection Set 2

B: Viability of cells at time points: Trypan blue exclusion

Figure 3.10 ICC γH2AX Damage Profile 2 of MRC5 and XP14 following MLV transduction. A repeat of the immunocytochemistry analysis seen in **figure 3.9** with 100 cells. **A)** shows the number of γH2AX foci per cell which indicate the cumulative mean of DSBs in a collective amount of cells of the MRC5 and XP14 cell lines. **B)** shows the cell viabilities of both cells lines staying above 90% for all time points for both cell lines. Consistent titres were used on each cell line. Error bars represent ±SEM.

The results seen in the replicate experiment seen in **figure 3.10** are similar to the ones seen in original experiment seen in **figure 3.9**. An initial low DSB count following DSB numbers peaking at the 6 hour mark. This was followed by a reduction in DSBs in both cell lines with a more pronounced reduction in DSBs in the MRC5 cell line, the XP14 cell line showed a similar plateau in DSB numbers from the 12-48 hour time points.

3.4.2 PA317 Cell Line Produced Empty MLV Infections

The next set of infections were performed using genome lacking "empty" MLV procured from the PA317 mouse cell line lacking a genome as described in the materials and methods **section 2.2.5**. Consistent titres were used on each cell line.

3.4.2.1 Empty MLV Infection Set 1
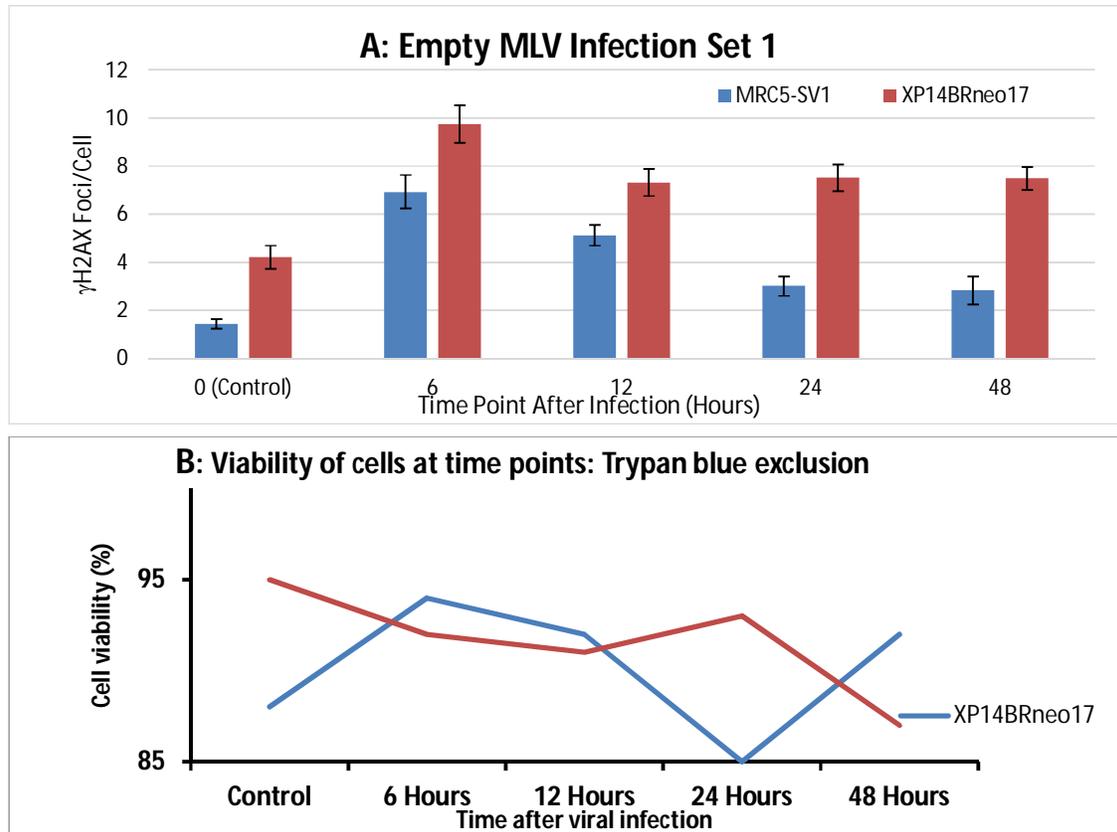
The results seen in **figure 3.11** are similar to the ones seen in **figure 3.9** and **3.10**. With low DSB numbers at the 0 hour control time point, followed by a peak at the 6 hour time point for both cell lines and a gradual reduction up to the 48 hour mark for the MRC5 line and a plateau for the XP14 line from the 12 hour mark. This shows that the viral particles lacking a genome can also cause similar damage to viral particles with a genome.

3.4.2.2 Empty MLV Infection Set 2
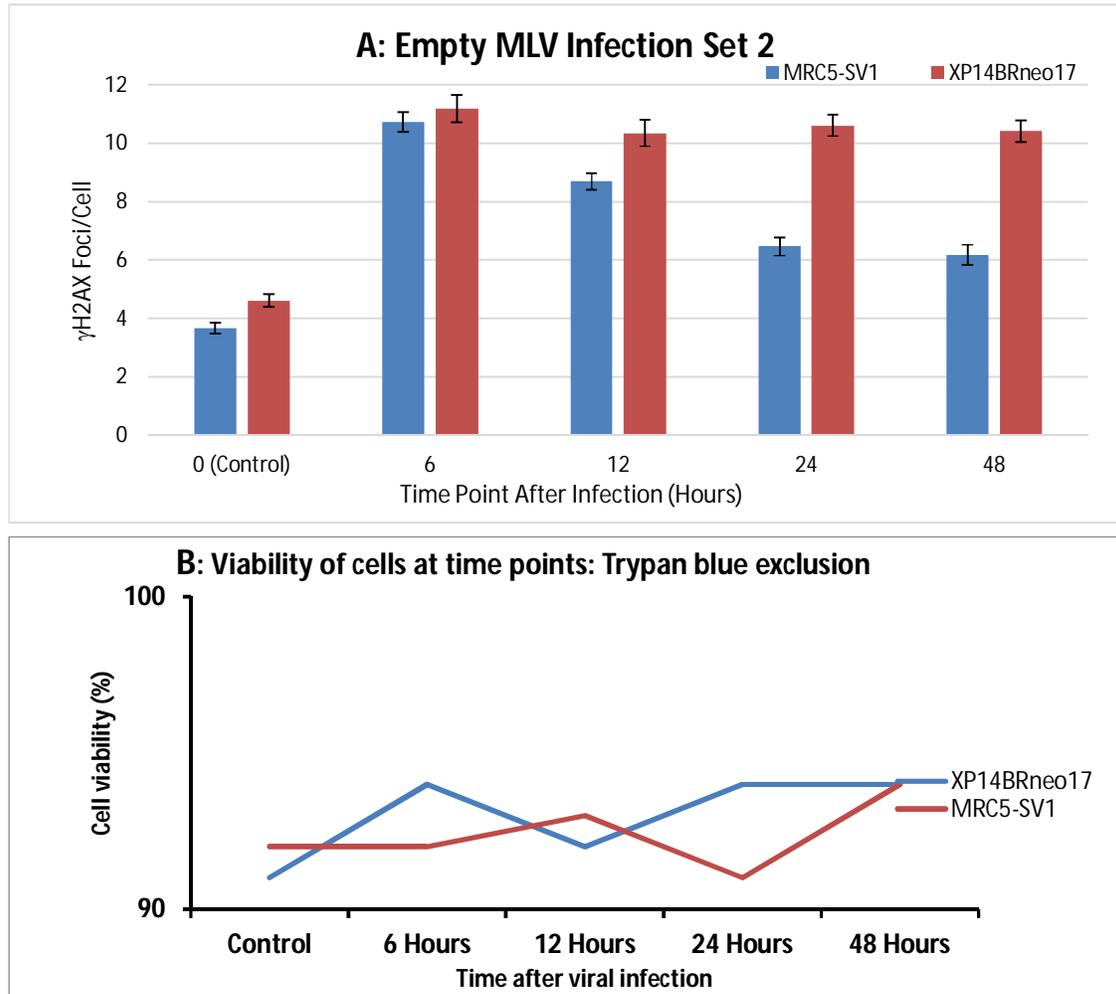
Figure 3.12 ICC γH2AX Damage Profile 2 of MRC5 and XP14 following empty MLV transduction.
A repeat of the immunocytochemistry analysis seen in **figure 3.11** with 100 cells. **A**) shows the number

of γH2AX foci per cell which indicate the mean of DSBs in a collective amount of cells of the MRC5 and XP14 cell lines. **B**) shows the cell viabilities of both cells lines staying above 90% for all time points for both cell lines. Consistent titres were used on each cell line. Error bars represent ±SEM.

The results seen in **figure 3.12** are similar to the ones seen with the empty viral infections in **figure 3.11** as well as the viral infections with a genome seen in figures **3.9** to **3.10**. Both cell lines show a low DSB count at the 0 hour control time point followed by a peak at the 6 hour time point. The DNA repair competent MRC5 cell line gradually repairs up to the 24 hour time point before plateauing while the DNA repair impaired XP14 cell line plateaus at the 12 hour time point.

3.4.2.3 Empty MLV Infection Set 3
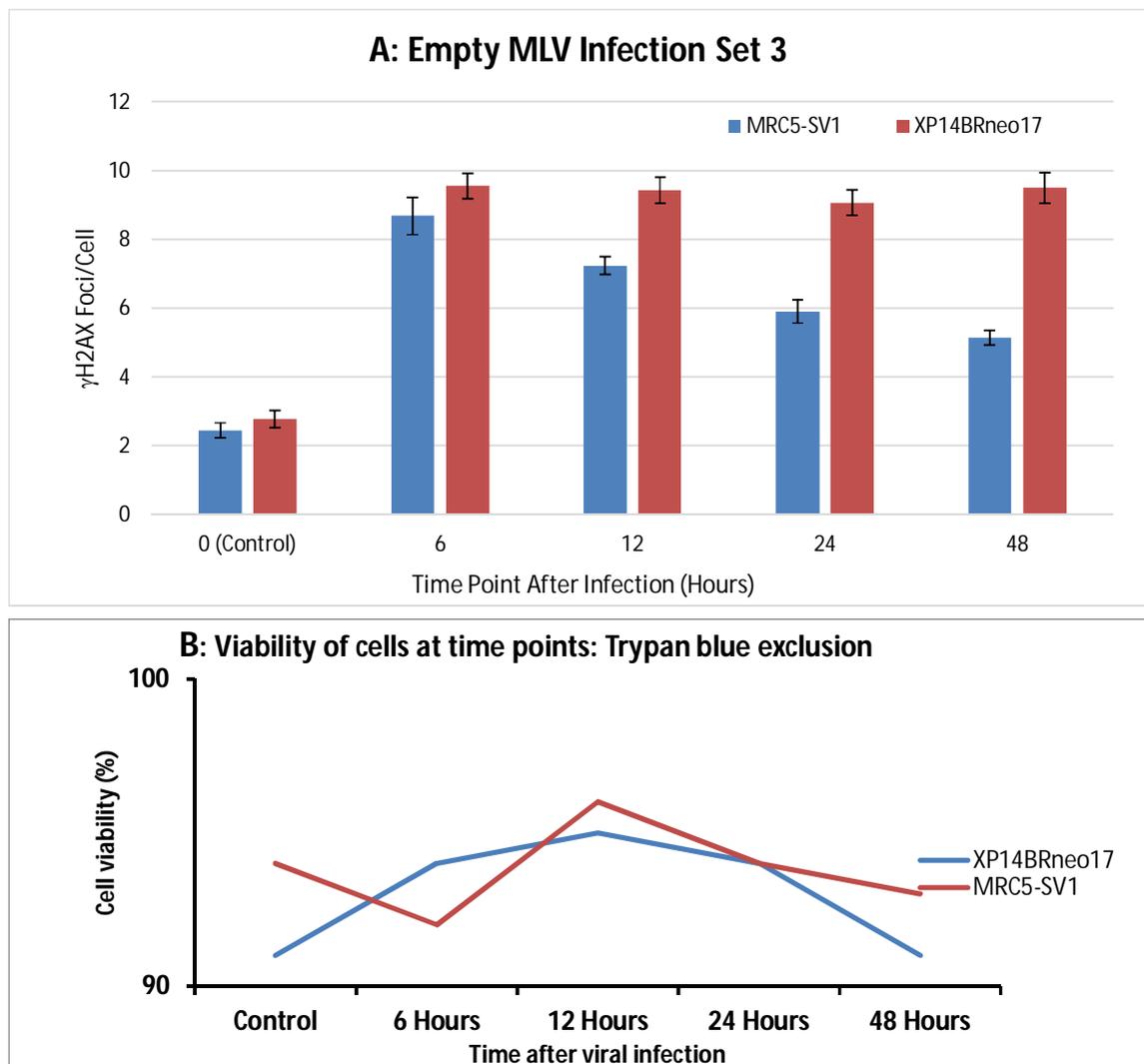
Figure 3.13 ICC γH2AX Damage Profile 3 of MRC5 and XP14 following empty MLV transduction.
A repeat of the immunocytochemistry analysis seen in **figure 3.11** and **3.12** with 100 cells. **A**) shows

The results seen in **figure 3.13** are similar to the ones seen with the empty viral infections seen in **figures 3.11** and **3.12**. All three experiments are replicates of each other and indicate similar results. With this repeat, the XP14 line shows no repair, DSBs peaking at the 6 hour mark and plateauing.

The three repeats done with "empty" MLV indicated less DSB repair seen in the NHEJ impaired XP14 cell line as compared to the NHEJ-competent MRC5 cell line.

3.4.3 Complete HIV Vector Aliquot Infections

So far experiments were done with an unknown consistent titre of virus obtained from the supernatant of viral producer cells. The experiments were then repeated with HIV vectors provided by Prof. Mike Themis with known titres at an MOI of 10 as described in the materials and methods **section 2.2.5**.

3.4.3.1 Complete HIV Infection Set 1
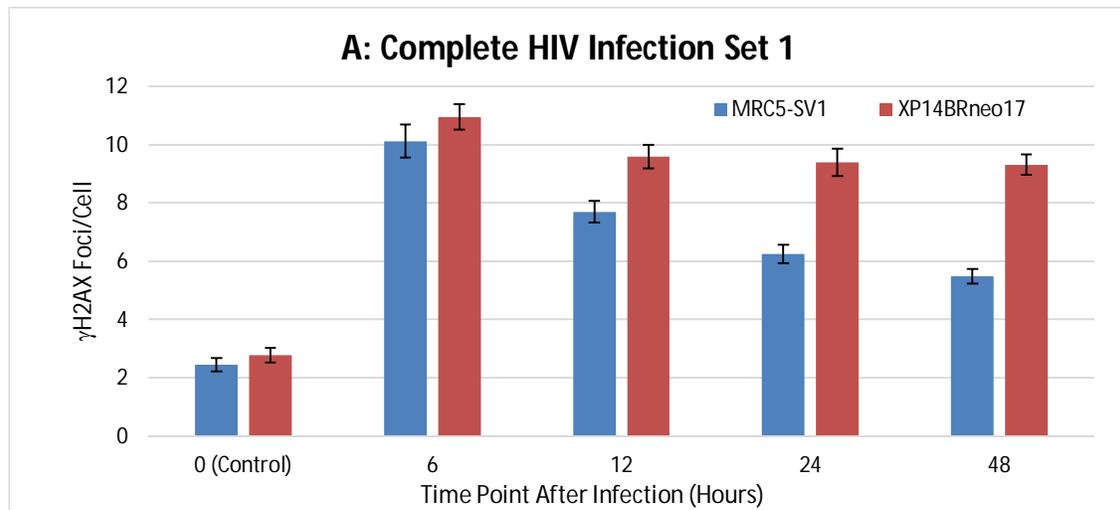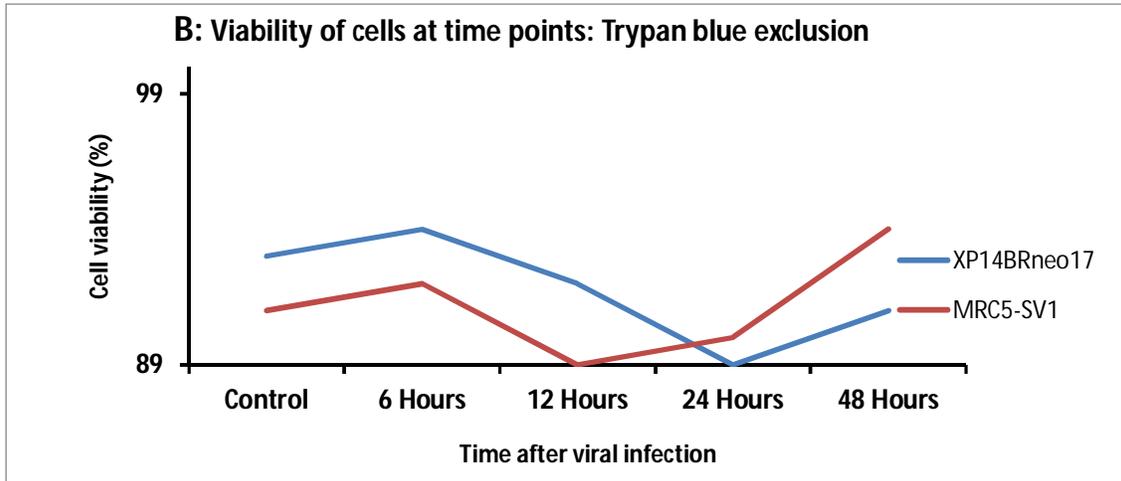
Figure 3.14 ICC γH2AX Damage Profile 1 of MRC5 and XP14 following HIV LV transduction. Immunocytochemistry analysis of 100 cells following infection and fixed at known time points. **A**) shows the number of γH2AX foci per cell which indicate the mean of DSBs in a collective amount of cells of the MRC5 and XP14 cell lines. **B**) shows the cell viabilities of both cells lines staying above 89% for all time points for both cell lines. Infections were done at an MOI of 10. Error bars represent ±SEM.

The results seen in **figure 3.14** are similar to the ones seen with the MLV infections done in **section 3.4.2**. The general pattern of the DSB counts remains, with a 0 hour control time point having the lowest number of DSBs which then peaks at the 6 hour point after infection. The MRC5 cell line shows repair down to the 48 hour mark while the XP14 line shows a small repair up to the 12 hour mark followed by a plateau.

3.4.3.2 Complete HIV Infection Set 2
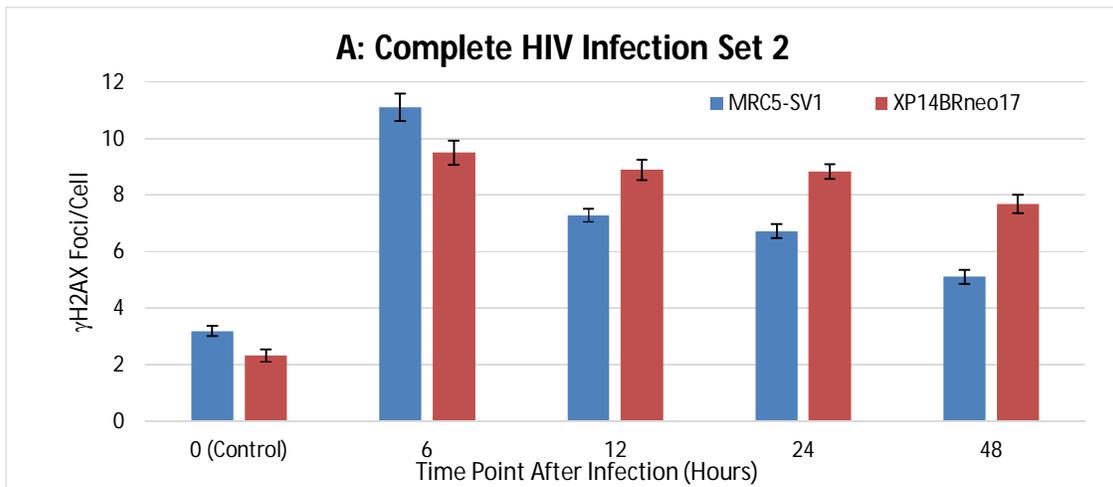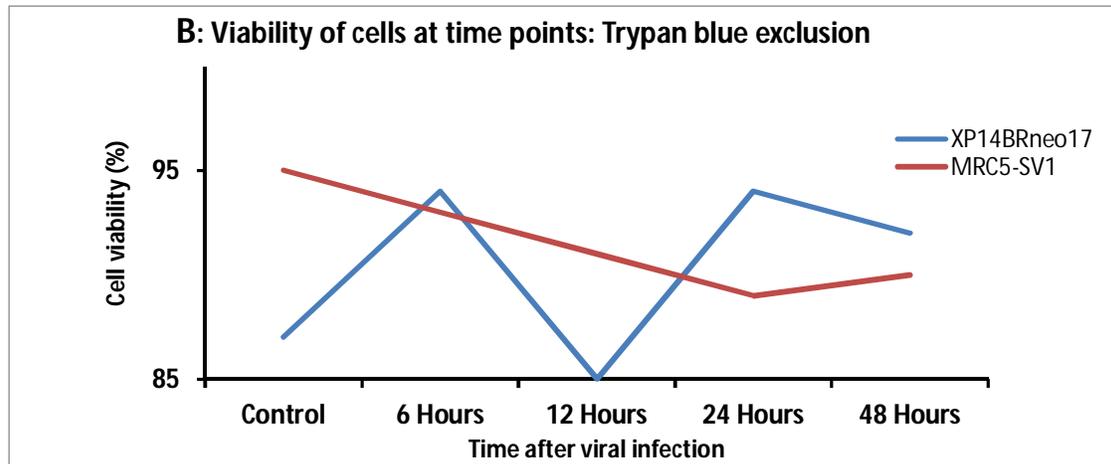
B: Viability of cells at time points: Trypan blue exclusion

Figure 3.15 ICC γH2AX Damage Profile 2 of MRC5 and XP14 following HIV LV transduction. A repeat of the immunocytochemistry analysis seen in **figure 3.15** with 100 cells. **A**) shows the number of γH2AX foci per cell which indicate the mean of DSBs in a collective amount of cells of the MRC5 and XP14 cell lines. **B**) shows the cell viabilities of both cells lines staying above 85% for all time points for both cell lines. Error bars represent ±SEM.

The results seen in **figure 3.15** are similar to those seen in **figure 3.14**. Both experiments are replicates. Both cell lines indicate their lowest DSB count at the 0 hour control time point and both peak at the 6 hour time point after infection. The MRC line repairs damage all the way to the 48 hour time point while the XP14 repairs up to the 12 hour point and plateaus up to the 24 hour mark after which there is a further reduction in DSBs at the 48 hour mark.

In these two cases, the repair competent MRC5 line shows more repair and a general lower amount of DSBs formed compared to the NHEJ incompetent XP14 line. These results following infections with known titres are consistent with the supernatant MLV infections with unknown titres seen in **section 3.4.2**.

3.4.4 Empty HIV Vector Aliquot Infections

The next set of infections were performed using "empty" HIV vectors lacking genome at a known titration. These aliquots were provided by Prof. Mike Themis. Infections were done at an MOI of 10 as described in the materials and methods **section 2.2.5**.

3.4.4.1 Empty HIV Infection Set 1
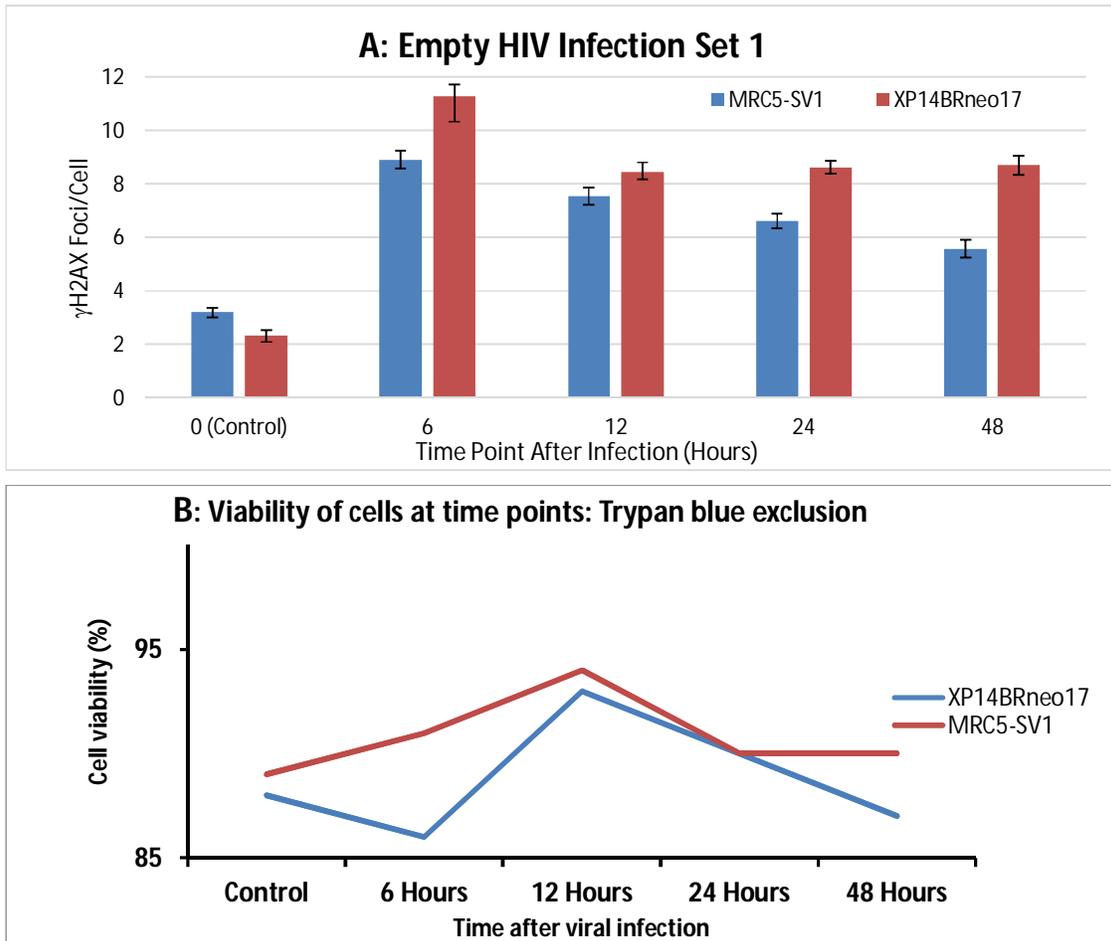
Figure 3.16 ICC γH2AX Damage Profile 1 of MRC5 and XP14 following empty HIV LV transduction. Immunocytochemistry analysis of 100 cells following infection and fixed at known time points. **A**) shows the number of γH2AX foci per cell which indicate the mean of DSBs in a collective amount of cells of the MRC5 and XP14 cell lines. **B**) shows the cell viabilities of both cells lines staying above 85% for all time points for both cell lines. Infections were done at an MOI of 10. Error bars represent ±SEM.

The results seen in **figure 3.16** are similar to those seen in **figure 3.14**. Both cell lines have the lowest DSB count at the 0 hour control time point followed by a peak at the 6 hour time point. MRC shows repair up to the 48 hour mark while XP14 plateaus at the 12 hour mark.

3.4.4.2 Empty HIV Infection Set 2
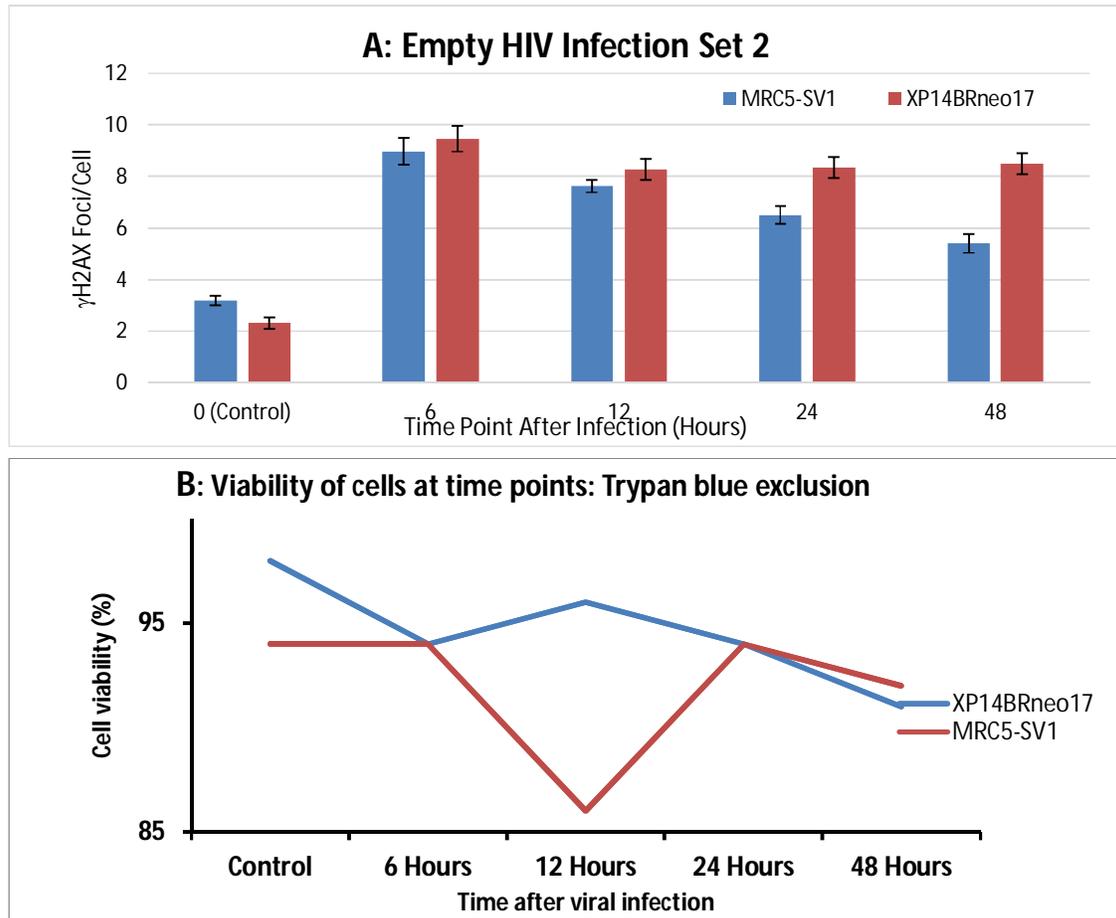
Figure 3.17 ICC γH2AX Damage Profile 2 of MRC5 and XP14 following empty HIV LV transduction. A repeat of the immunocytochemistry analysis seen in **figure 3.16** with 100 cells. **A**) shows the number of γH2AX foci per cell which indicate the mean of DSBs in a collective amount of cells of the MRC5 and XP14 cell lines. **B**) shows the cell viabilities of both cells lines staying above 85% for all time points for both cell lines. Error bars represent ±SEM.

A similar profile is seen in **figure 3.17** as that seen in **figure 3.16**. The two repeats done with empty HIV vectors indicated less DSB repair seen in the NHEJ impaired XP14 cell line as compared to the NHEJ-competent MRC5 cell line. The lowest DSB count being at the 0 hour control time point followed by a peak at the 6 hour time point. MRC shows repair up to the 48 hour mark while XP14 plateaus at the 12 hour mark.

3.5 Discussion

3.5.1 General Results

An assay to detect γH2AX (a marker of DSBs) was utilised by applying fluorescent marker coupled antibodies which bind to γH2AX. ImageStream and Immunocytochemistry were used to detect these markers.

3.5.1.1 ImageStream Analysis of DSB Profiles

The ImageStream analysis results indicated high background DNA damage as seen in **section 3.3**. At the 0 hour control time point, the ImageStream results indicated a high amount of DSBs compared to 6 hours following viral infection across all analyses. This would mean that, at a base level, without any viral agent to cause DNA damage, cells have a greater amount of DSBs which are then reduced once a viral infection occurs.

Those results were at odds with the prior work done as shown in **appendix 1** and **2** where numbers of DSBs are lowest at the 0 hour mark and peak at the 6 hour mark after infection.

This has also been shown with irradiation damage applied to cells to cause DNA damage which was detected via γH2AX assays. In these experiments, the lowest amount of γH2AX foci, or DSBs, was at the 0 hour time point followed by an increase in DSB formation (Redon *et al.*, 2009; Lee *et al.*, 2019). Published research utilising the γH2AX assay has also shown that viral infections can cause DSBs which are at their lowest level at the 0 hour time point before increasing (Skalka and Katz, 2005; Tarakanova *et al.*, 2007; Botting, Lu and Triezenberg, 2015).

The incongruence with prior work and published research indicates that an error had occurred at some point in the ImageStream method. This could have been during the transduction phase, the antibody application phase or the ImageStream data analysis phase. In order to determine if ImageStream data analysis was being performed incorrectly, the raw data obtained from work prior to these experiments which provided the results seen in **appendix 2**. The reanalysis results in **figure 3.6** and **3.7** are inconsistent from the prior results seen in **appendix 2**. The inconsistency is most likely due to an error during the analysis. One possible area of error during analysis is gating. Gating is required to filter out debris and clumps of cells from data to be analysed. For this, the area that is most likely to be single cells is defined for the analysis. Only material which falls into this area is considered for γH2AX foci counts. It is possible, that

the gating is picking up too many small debris, leading to more foci counts than actually present in single cells.

3.5.1.2 Immunocytochemistry Analysis of DSB Profiles

The results seen in **section 3.4** in **figures 3.9 to 3.17** match the trends seen in the prior data seen in **appendix 1** and **2** as well as prior published data (Skalka and Katz, 2005; Tarakanova *et al.*, 2007; Botting, Lu and Triezenberg, 2015).

In both repair competent, MRC5, and NHEJ impaired, XP14, cell lines, a low number of DSBs at the 0 hour control point, followed by an increase in DSB damage at the 6 hour mark. This was followed by a period of significant repair in the NHEJ competent MRC5 cell line by the 48 hour point, but little significant repair in the NHEJ compromised XP14 cell line by the 48 hour point. However, even in the DNA damage repair competent cell line, the level of DSBs do not return to the levels they were at the 0 hour control time point, indicating that not enough time has passed to allow for complete repair, or that complete repair is not possible following viral infection. Replicate experiments gave similar results, indicating the reliability of the results.

DSB formation is required for the retroviral life cycle in the integration stage. It was possible to confirm the integration step for complete vectors to ensure that it was integration that was causing the DSB formation. The "complete" vectors contained the *GFP* gene which, when expressed, will produce the GFP protein in cells which will glow green under 488 nm wavelength light. Following complete vector integration, the glow of the GFP could be observed at the 6 hour mark to confirm integration, in all cells which would appear as follows with successful integration.
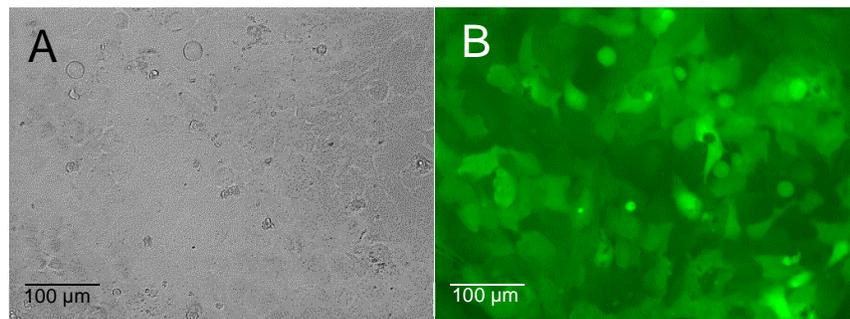
Figure 3.18 Green Fluorescence Indicating Integration. A) Shows MRC5 cells under brightfield microscopy following GFP retroviral vector infection. **B**) Shows the same cells under 488 nm wavelength light, which causes them to glow green confirming successful integration has taken place. Scale Bars = 100µm.

This method was not able to be used to show "empty" vector integration as no *GFP* gene was present in the vector to integrate into the host cells. As such the comparable pattern of DSB profiles between empty and complete vectors being similar could indicate a similar integration process occurring with the "empty" vectors.

The results indicate that DNA undergoes DSB damage following infection. Especially so in the NHEJ incompetent XP14 cell line. This means that candidates for gene therapy utilising retroviral vectors should be evaluated for NHEJ competence before the use of viral gene therapy to better advise them.

These results have not been shown in prior publications, as most DNA damage assays utilising γH2AX assays are based on DNA damage caused by irradiation (Redon *et al.*, 2009; Lee *et al.*, 2019) or for early detection of carcinogenesis in rats (Toyoda and Ogawa, 2022; Toyoda *et al.*, 2023). As such, retroviral vector effect on DSB formation is not published on heavily, most such studies being used as markers to identify the infection by viruses such as herpesviridae (Tarakanova *et al.*, 2007; Botting, Lu and Triezenberg, 2015) rather than the effects of retroviral vectors utilised for gene therapy.

3.5.2 Further Work

More time points beyond the 48 hour point could be assessed using ICC for the MRC5 line to determine if repair occurs at a later time. Refinement of the ImageStream data analysis methodology could corroborate findings from the ICC experiments.

Other assays to detect DSBs could also be considered, such as the comet assay and immunoblotting. The comet assay is performed by first encapsulating cells in agarose gel on a microscope slide, followed by which the cells are lysed using chemical agents. Then electrophoresis is performed and, depending on the amount of DSBs, a trail will be left in the agarose left by the DNA of the cells. The longer the trail, the more the DSB damage has occurred (Nandhakumar *et al.*, 2011). Immunoblotting, such as Western Blotting, would involve the staining of the γH2AX proteins with fluorescent coupled antibodies in cells and using gel electrophoresis to determine the density of fluorescence to provide a semi-quantitative assay to determine the amount of γH2AX proteins formed (Toyooka, Ishihama and Ibuki, 2011). Vectors produced for gene therapy could be investigated to find out if empty vectors are also packaged alongside them. These vectors can reduce the purity and efficacy of the batch, and may also cause DNA damage in the form of DSB formation in gene therapy patients, especially those with impaired NHEJ mechanisms.

Chapter 4 – Genotoxicity Assays of Cells Infected with Complete and Empty Retroviral Vectors

4.1 Background and Aims

As mentioned in **chapter 1** and **3**, retroviral vectors for gene therapy function by making breaks in the target cell DNA and inserting a therapeutic DNA sequence into the gaps formed by the breaks (Skalka and Katz, 2005).

There is a possibility that the breaks formed can cause genotoxic damage to the target cell DNA by potentially disrupting existing genes within the cells. Genotoxicity is a property of a chemical agent to induce an uncontrolled change in the genetic material of a cell and leads to a mutagenic change in phenotype such as cancer (Ramezani, Hawley and Hawley, 2008; Biasco *et al.*, 2017; Morgan *et al.*, 2021). Retroviral vectors directly alter the genes of a host cell and as such could cause unintended damage to the host DNA via genotoxicity. The HPRT assay can be used to determine the potential genotoxic effects of these vectors. Gene knockout by retroviral vectors has been demonstrated in the past utilising the *HPRT* assay (Themis *et al.*, 2003; Grosovski *et al.*, 1993; King *et al.*, 1985; Bradley *et al.*, 1981).

The assay works in a two-step selection process to determine mutagenesis in the *HPRT* gene. The gene is targeted for its location on the X chromosome, thus allowing for a single gene knock-out assay in male cells which only have a single X chromosome. The gene is also targeted for its role in the salvage pathway of DNA synthesis in the absence of the *de novo* pathway (Åkerlund *et al.*, 2017, Themis *et al.*, 2003; Grosovski *et al.*, 1993).

4.1.1 Aims and Objectives

- To select male cell lines with and without NHEJ pathways in HAT medium as part of the first selection stage of the HPRT assay

- Treat cells with virus infection to show their mutagenesis at the *HPRT* locus

- To determine whether cell infection by virus carrying genome or no genome cause *HPRT* mutagenesis via a second selection process in either 6TG or 8AG

- To determine the above where cells do not have an intact NHEJ pathway

The first step of selection involves growing cells in HAT medium which will only allow cells with the HPRT gene to survive. Following this selection, the genotoxic agents (the retroviral vectors) to be tested will be applied to the cells. Then the second selection process is done by growing the cells in 6-thioguanine (TG) or 8-azaguanine (AZ), which will only allow cells with impaired *HPRT* genes to survive, meaning that all the cells that survive the second selection have most likely had their *HPRT* gene rendered non-functional due to the genotoxic effect of the viral vector used. This allows for a quantifiable method to determine the genotoxicity of each vector, as the more cells that survive the second selection process, the more genotoxic the viral vector used was.

This two-step selection process was planned to be used on male NHEJ competent and incompetent cell lines with "empty" and "complete" vectors used to infect the host cells in between selections to determine the genotoxic potential of the vectors.

4.2 HPRT Assay

The HPRT assay works in a two-step selection process to determine genotoxicity. The *HPRT* gene is used in the salvage pathway of cells for the synthesis of DNA, which is required in the absence of the *de novo* synthesis pathway. The gene is found on the X chromosome, thus only one copy will be present in male cells with XY chromosomes.

This makes the gene a good candidate for genotoxicity assays, as only one copy in male cells needs to be disrupted in order to indicate the presence of genotoxicity as opposed to genes not found on the X chromosome where potentially two copies would have to be disrupted in order to show genotoxic effects.

The first selection is done in HAT (hypoxanthine-aminopterin-thymidine) medium. Aminopterin is a drug which blocks the *de novo* synthesis of nucleotides forcing the cells growing to utilise the salvage pathway. The salvage pathway uses hypoxanthine and thymidine to form the nucleotides required for DNA synthesis. HPRT is used to process hypoxanthine in this process. HAT medium can thus be used to select a population of cells which have the functional *HPRT* gene as the cells lacking functional *HPRT* will be unable to use the salvage pathway to generate new nucleotides, which leads to cell death as seen below in **figure 4.1** (Åkerlund *et al.*, 2017; Themis *et al.*, 2003; King *et al.*, 1985; Grosovski *et al.*, 1993).
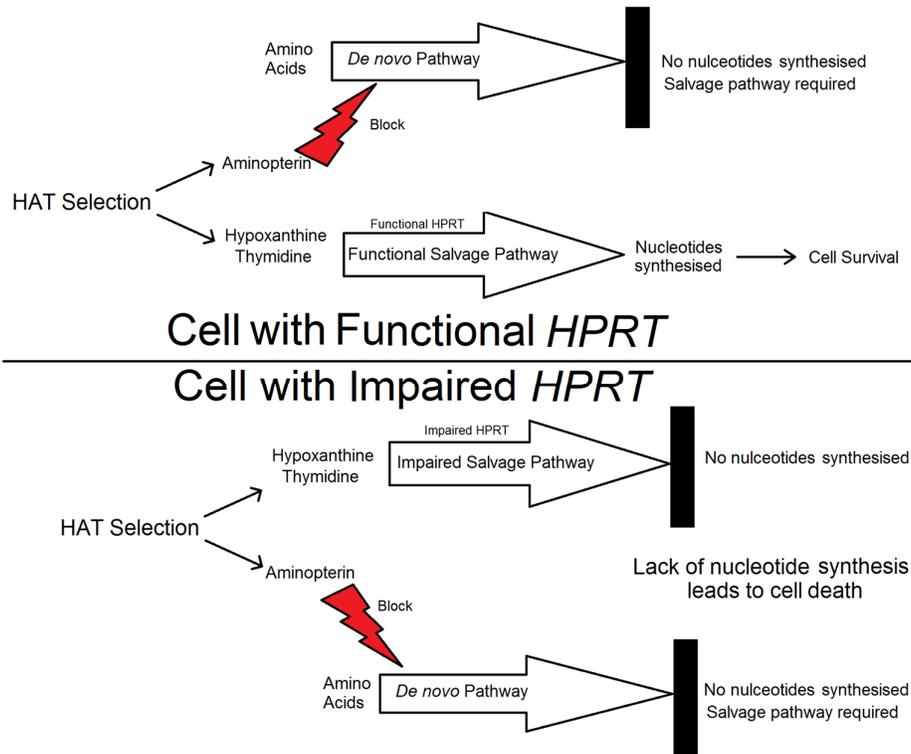
Figure 4.1 HAT Selection Principle. HAT medium contains hypoxanthine, aminopterin and thymidine. Aminopterin will block the *de novo* nucleotide synthesis pathway forcing the cell to utilize the salvage pathway. The medium provides hypoxanthine and thymidine, which are critical for the salvage pathway and allow the cells to still produce nucleotides. The salvage pathway utilizes HPRT to process hypoxanthine in the salvage pathway. Thus, if the *HPRT* gene is already non-functional, the cell will die due to lack of nucleotides. Thus HAT medium can select for cells which have a functional *HPRT* gene (Åkerlund *et al*., 2017; Themis *et al*., 2003; King *et al*., 1985; Grosovski *et al*., 1993). Image made using Microsoft Paint.

Following HAT selection, all surviving cells will have functional *HPRT*. These cells can then be exposed to genotoxic agents to determine how genotoxic they are. For this thesis, the viral vectors will be tested as they were in **chapter 3**. The genotoxic agents can then possibly damage *HPRT* functionality, and the more genotoxic the agent, the more the cells that will be affected.

A second selection can then be used utilising the toxic purine analogues, 6-thioguanine (6TG) or 8-azaguanine (8AG). Cells with functional *HPRT* will process these in the DNA salvage pathway to form toxic compounds which will cause the cells to die. Thus, only cells with non-functional *HPRT* will be able to survive this selection by using the *de novo* pathway.
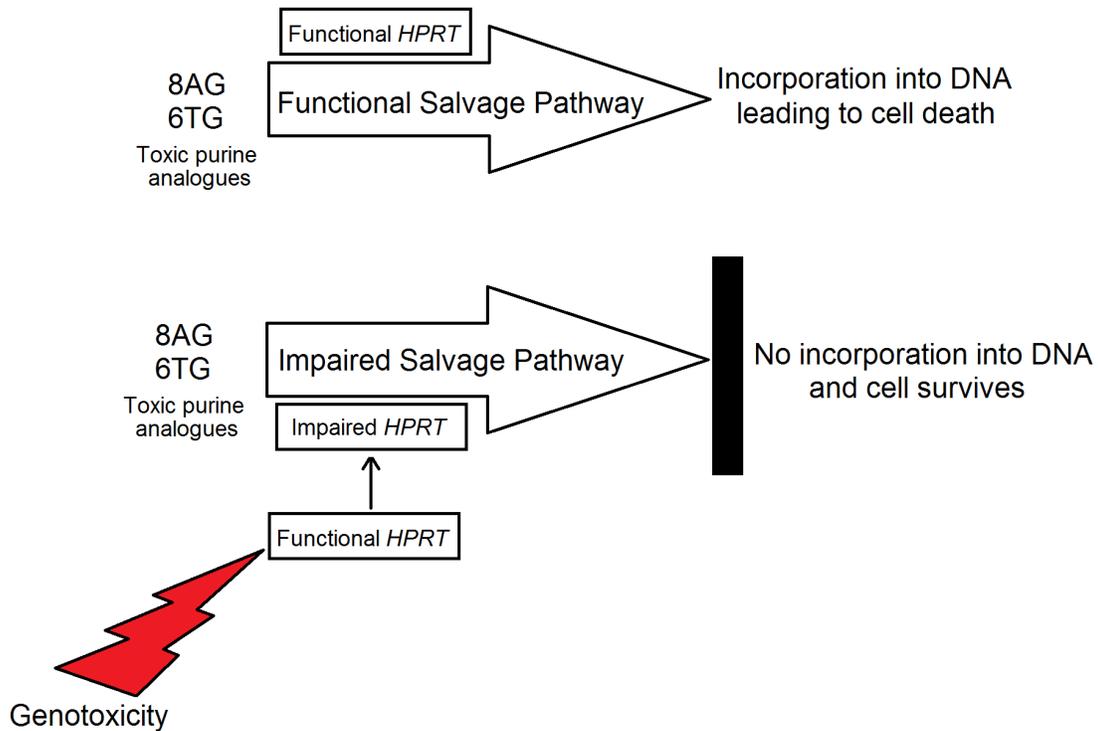
Figure 4.2 *HPRT* Assay Principle. After the application of elements which cause genotoxicity, such as viral infection, the *HPRT* gene may become mutated. The rate of mutation can be determined by applying 8AG or 6TG to the cells. HPRT will process 8AG or 6TG into toxic purine analogues which will then be incorporated into cell DNA. These cells would then die. However, cells with mutant *HPRT* will be unable to use the salvage pathway, and thus be unable to incorporate 8AG and 6TG into the DNA, thus the cell survives (Åkerlund *et al*., 2017; Themis *et al*., 2003; Grosovski *et al*., 1993; King *et al*., 1985 Bradley *et al*., 1981). Image made using Microsoft Paint.

A clear two-selection process emerges which can be used to test genotoxic profiles. A population of cells selected using HAT medium with functional *HPRT* can be infected with the vector being tested which will lead to the disruption of the *HPRT* gene in a certain number of cells. The number of cells affected will be proportional to the genotoxicity of the vector. A following selection in either 6TG or 8AG will then select a cells which do not have functional *HPRT*. A simple count of the colonies following this second selection provides a number which corresponding to the genotoxicity of the *HPRT*, a higher number of colonies indicating a higher degree of genotoxicity (Åkerlund *et al*., 2017; OECD, 2016; Themis *et al*., 2003; King *et al*., 2007; Bradley *et al*., 1981).

Male cell lines were selected for the *HPRT* assay, as the *HPRT* gene is present on the X chromosome, ensuring male cells only have one copy, and thus only a single gene needs to be knocked out to show an effect.

For this reason, the male cell lines with functional DNA repair mechanisms, V79 and MRC5SV1, along with the DNA repair impaired male cell line, AT5BIVA, were selected to determine the role of DNA repair mechanisms in preventing genotoxicity.

Then, following HAT selection, the cells would be infected with the retroviral vectors tested in **chapter 3** for DSB formation. This allows us to get a better picture of the extent of harm that retroviral vectors could cause by also getting their genotoxic profiles.

4.3 HAT Selection of Male Cell Lines

As part of the first selection for the *HPRT* assay, the male cell lines V79, MRC5SV1 and AT5BIVA were grown in HAT medium for five days to remove all cells that had lost their ability to utilise the HPRT salvage pathway. Hence, any cells lacking *HPRT* function following this selection are more likely to have lost this function due to genotoxicity caused by retroviral vector function at the *HPRT* locus. V79 and MRC5 cell lines are both DNA repair competent while AT5BIVA is not. HAT selection was done with the method described in the **materials and methods section 2.2.12**.

MRC5SV1
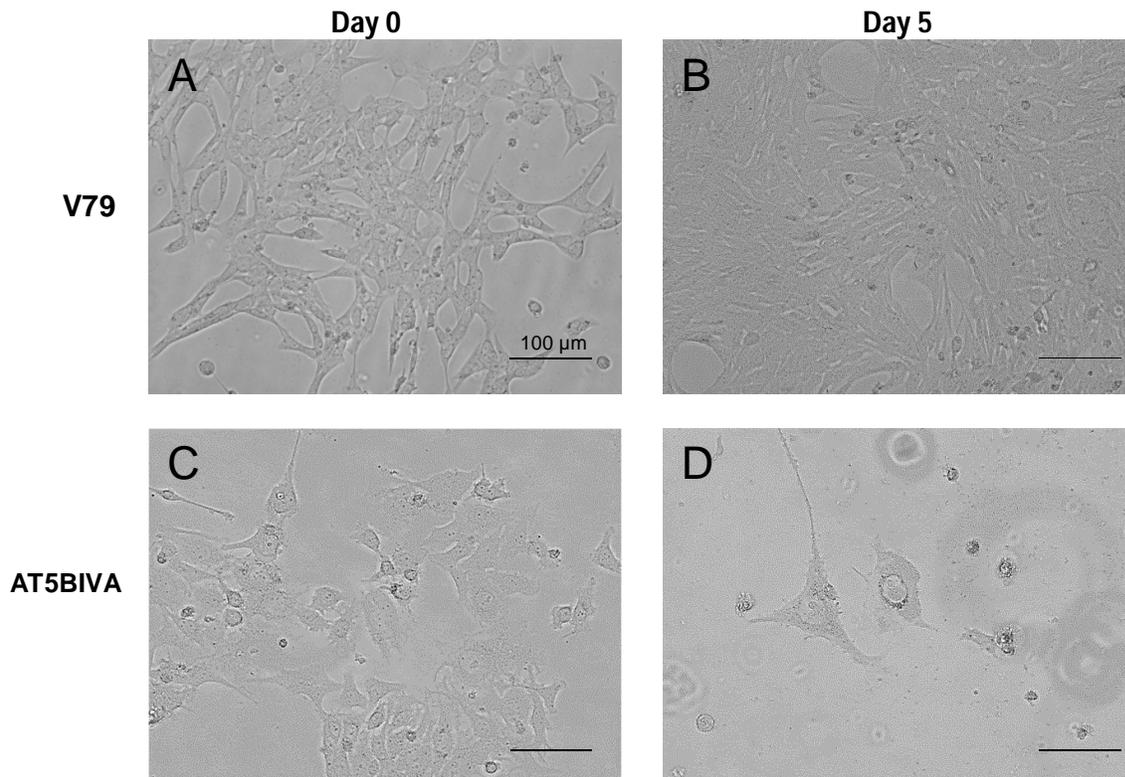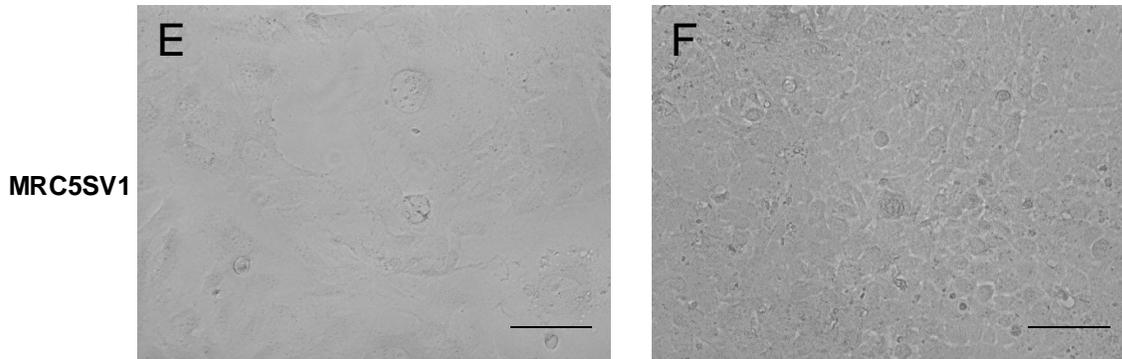
Figure 4.3 V79, AT5 and MRC5, grown in HAT medium over 5 days. Figures **A, B** are V79 cells, **C, D** are AT5 cells and **E, F** are MRC5 cells. The cell lines were grown to 90% confluency before passage as was required. The cell lines are shown at the start of HAT selection (Day 0) (**A, C, and E**) and after 5 days of HAT selection (**B, D and F**). HAT selection was done with the method described in the materials and methods **section 2.2.12**. The photomicrographs were taken at 20x magnification with an EVOS FLoid Imaging System. Scale bars = 100 µm.

Figure 4.3 shows the MRC, V79 and AT5 cell lines as they grew in HAT medium. In order to compensate for a potentially low amount of innate genotoxicity (Themis *et al.*, 2003; Grosovski *et al.*, 1993; King *et al.*, 1985; Bradley *et al.*, 1981), 1x10^7 of each cell line was needed to be grown and exposed to the vectors. The following graphs outline the amount of cells grown over five days.
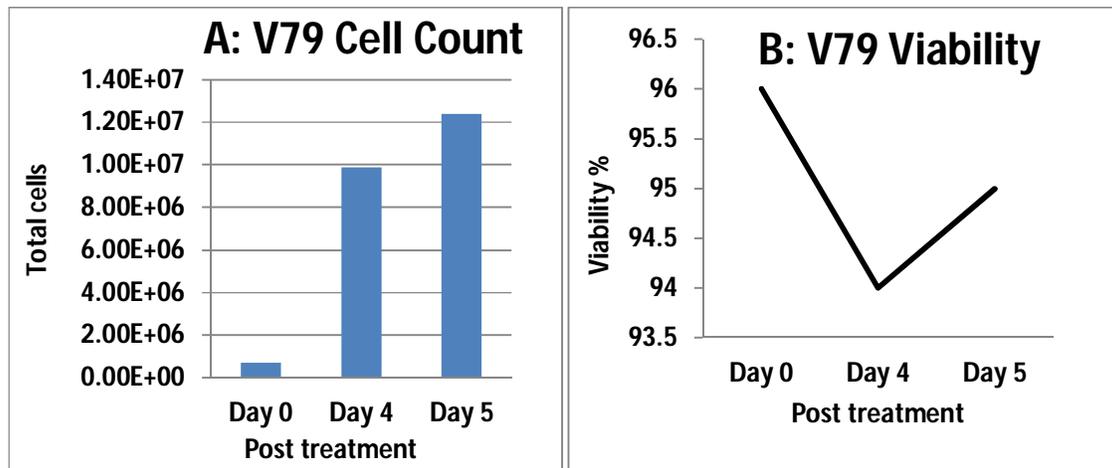
Figure 4.4 V79 cell numbers in HAT medium over five days. A) shows the progress of cell growth over 5 days. 7x10^5 V79 cells were seeded and grown in HAT medium and around 1.24x10^7 cells were able to be obtained by the fifth day. **B)** shows the viability of cells remaining over 94% over the entire period.

83
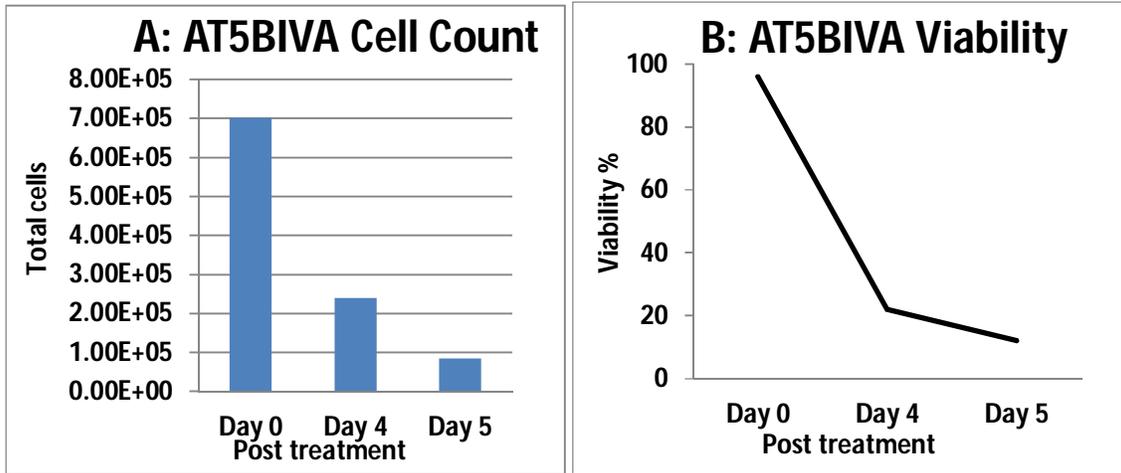
Figure 4.5 AT5BIVA cell numbers in HAT medium over five days. A) shows the progress of cell growth over 5 days. 7x10^5 AT5BIVA cells were seeded and grown in HAT medium**.** These cells did not grow well in the HAT medium and began dying in great numbers. By the 5th day only around 8.5x10^4 cells remained. **B**) shows the viability of cells dropping drastically from 96% to 12% by day 5.
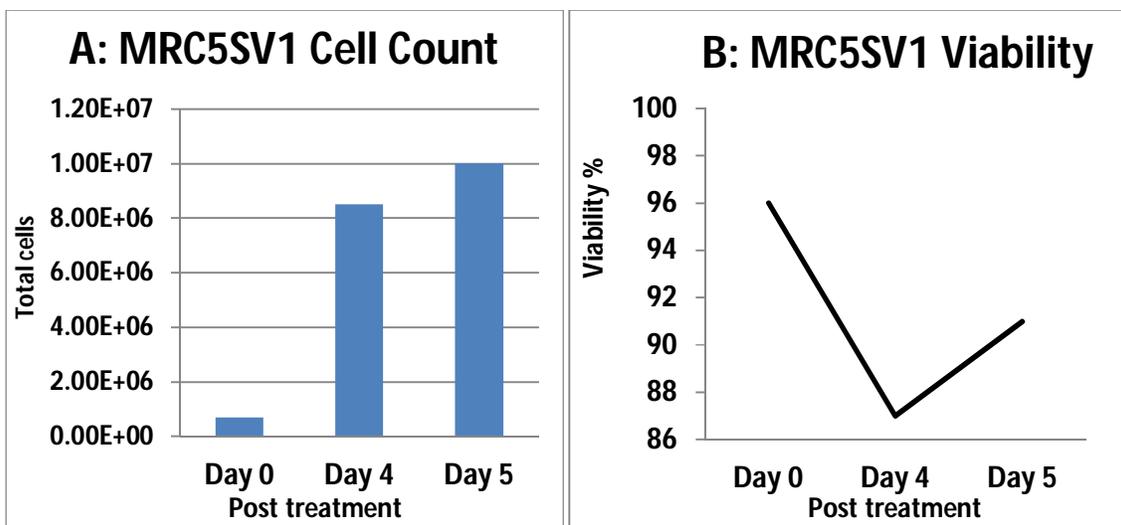
Figure 4.6 MRC5-SV1 cell numbers in HAT medium over five days. A) shows the progress of cell growth over 5 days. 7x10^5 MRC5-SV1 cells were seeded and grown in HAT medium. Around 1x10^7 cells were able to be obtained by the fifth day. **B**) shows the viability of cells remaining over 87% over the entire period.

As seen in **figures 4.4** and **4.6**, both respective repair competent cell lines, MRC5-SV1 and V79, were able to grow with a healthy viability for the five days up to the required 1x10^7 cells. These would then be infected with the viral vectors for the next phase of the study. As seen in **figure 4.5**, the AT5BIVA cell line was unable to grow optimally in HAT medium with poor viability and no growth. Repeat attempts to grown the AT5 line in

HAT medium yielded the same results. It is possible HAT selection is too intensive on a DNA repair impaired cell line.

4.4 GFP Vector Transduction Optimisation

To determine the genotoxicity of a retroviral vector, a demonstration of transduction and integration of the vector DNA into the host cell DNA can be used to show that retroviral integration could be responsible. One method to show successful integration following transduction is to use the *GFP* (*Green Fluorescent Protein*) gene in vectors used to infect the target cells. The *GFP* gene functions as a reporter of expression which, codes for the GFP protein which will glow green under ultraviolet light. Thus, the presence of a green glow in cells following transduction of the *GFP* gene under ultraviolet light shows successful transduction (Tsien, 1998; Phillips, 2001; Tian *et al.*, 2023).

MRC5-SV1, XP14BRneo17, V79 and AT5BIVA cells were transducted with *GFP* using retroviral vectors to determine if GFP expression could be detected prior to and following infection. The infections were done as described in **section 2.2.6**. The MRC5, V79 and AT5 cell lines were selected due for being female and thus the cells to be used for the *HPRT* assay. The XP14 line was used due to its availability prior to confirmation of being a female line. This expression was first optimised in non-HAT selected cells before use on the limited pool of cells that had undergone HAT selection. Once optimised, this method would be used to confirm successful "complete" retroviral transduction in cells.

4.4.1 Controls Showing Lack and Presence of Fluorescence

The cells were imaged using microscopy to demonstrate the lack of green florescence under the appropriate lighting. Flow cytometry was used to quantify the amount of green fluorescence seen in the untransducted cells; this was compared to the green fluorescence seen in a positive control in the form of a GFP lentivirus producer cell line which should show green fluorescence due to self-transduction. A cell line which was not part of the cell lines being investigates, 293T, was imaged and analysed as a negative control.
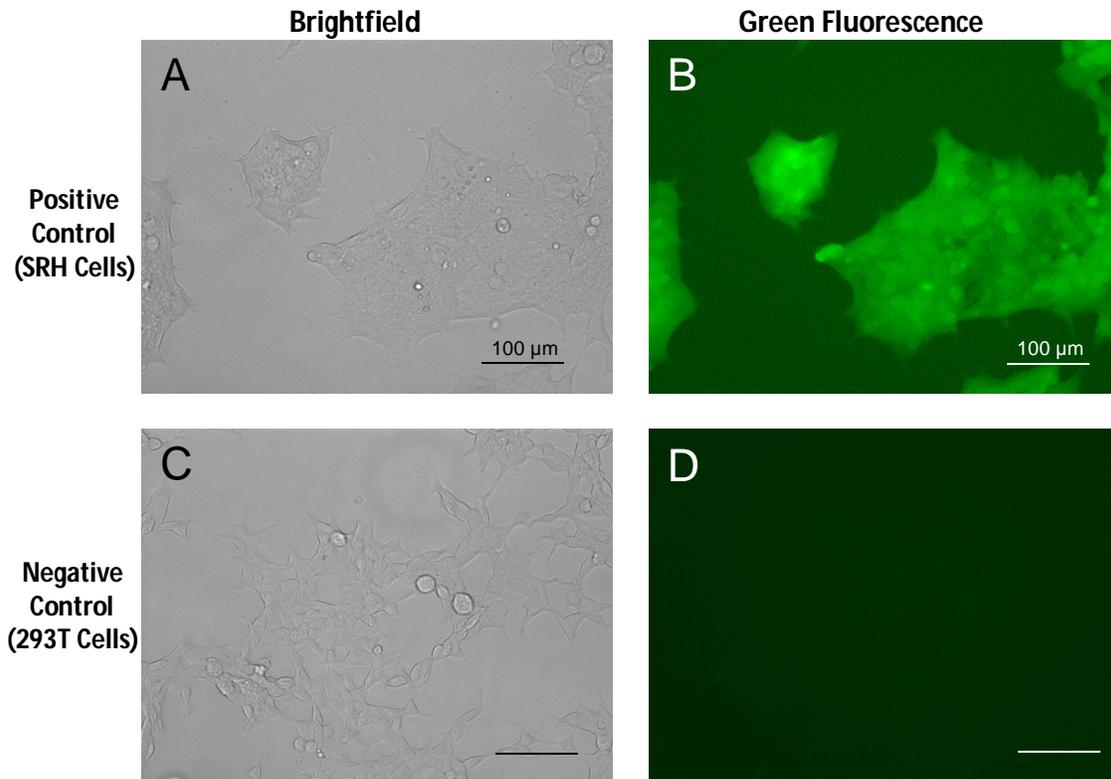
Figure 4.7 Control cell lines, SRH (+ve) and 293T (-ve) under brightfield and green fluorescence inducing light. The cells are shown with brightfield microscopy in **A**) and **C**) and fluorescent microscopy in **B**) and **D**). Green fluorescence is seen with the positive control cell line, SRH (**A** + **B**) while none is seen with the negative control cell line, 293T (**C** + **D**). The photomicrographs were taken at 20x magnification with an EVOS FLoid Imaging System. Scale bars = 100 μm.

Figure 4.7 shows how the negative control 293T cell line not glowing green due to a lack of any transfection with *GFP*. The positive control cell line, SRH, is a producer cell line which generates retroviral vectors containing the *GFP* gene. These vectors will infect the producer cells themselves, thus allowing the cells to make GFP protein which will then cause the cells to glow green under ultraviolet light. This makes producer cells a very intuitive positive control cell line for green fluorescence caused by *GFP* transfection. In order to quantify these visual results, flow cytometry can be utilised.

Flow cytometry can analyse and quantify a mass of trypsinised cells for fluorescence. However, a sample of trypsinised cells will also contain cell debris and clumped cells, both of which can produce erroneous results. Forward scattering (FSC) and side scattering (SSC) of light can be used to determine the population of cells within a sample of trypsinised cells. The greatest values of FSC and SSC of each object can be compared, referred to as FSC-H and SSC-H, H standing for the "height" or intensity of

the signal. Other values that can be measured are the FSC-A or SSC-A; the "A" standing for area, a description meaning the strength and duration of the scattering multiplied. FSC indicates the size of the object while SSC indicates the granularity. When compared to the FSC and SSC values of the entire population of objects being analysed in a sample, cell debris will have very small FSC values, due to their smaller size, and will have small SSC values, due to the lack of complex structures within. Clumped cells, on the other hand, compared to the general population, will have very high FSC values, due to the large size of the clumped together cells, and a large SSC value, due to the clumped cells' structures overlapping close to each other. As such, it is possible to "gate" a sample to only analyse what are most likely to only be cells, i.e. structures with moderate FSC and SSC values.

Once this gating is done, the individual cells can be quantified for green fluorescence. The FCS-H of the cells can be plotted against their fluorescein isothiocyanate (FITC)-H (the greatest signal of green fluorescence observed) or the FITC-A (the duration of fluorescence multiplied by strength. Cells which fluoresce green over a threshold can be considered to be glowing green. This allows for the analysis of the population of cells for a percentage of cells radiating green fluorescence.
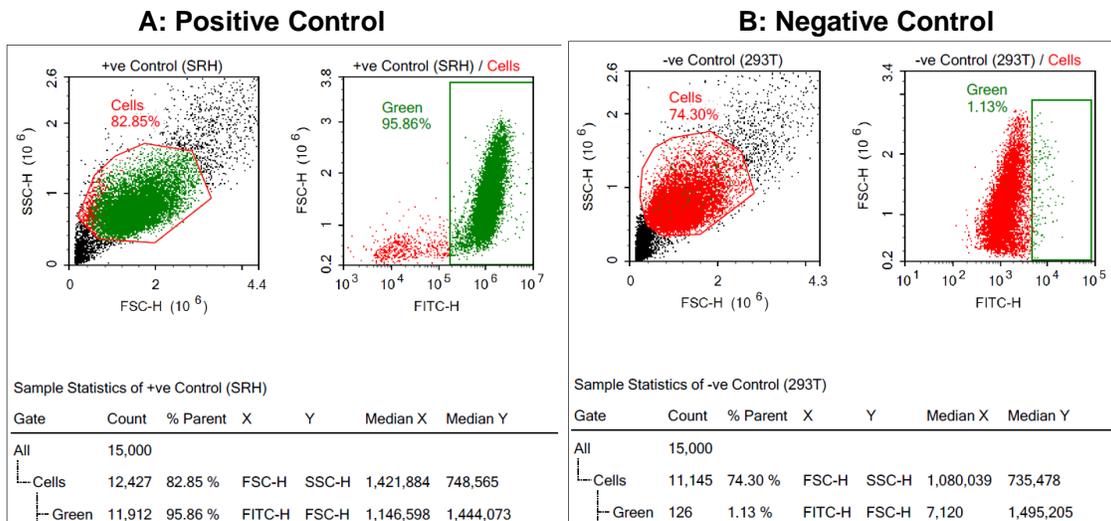
Figure 4.8 Flow Cytometry Results of Control Cell lines under green fluorescence inducing light. The graphs on the left are for the positive control, SRH cells. The graphs on the right are for the negative control, 293T cells. SSC-H and FSC-H were plotted to determine the population of cells in the sample and indicated as "cells". The FSC-H plotted against the FITC-H to determine the population of cells which were fluorescing green. The green rectangles indicate the area in which the cells would have to fall to count as glowing green. The left graphs (**A**) show the positive control results while the right graphs

The results in **figure 4.8** corroborate the visual results seen in **figure 4.7**. The positive control producer cell line had 95.86% of cells showing green fluorescence, while the negative control cell line showed only 1.13% of cells with green fluorescence. Thus, flow cytometry is able to quantify the amount of green fluorescence caused by *GFP* transfections.

The positive and negative controls can be used to determine the validity of further flow cytometry results of the test samples. The FITC-H value were the positive control was gated to show fluorescent cells can be used as a threshold value to indicate the necessary FITC-H signal required to indicate fluorescence in test samples. The test sample cell lines, MRC5-SV1, XP14BRneo17, V79 and AT5BIVA, were first imaged and analysed via flow cytometry to determine non-transducted fluorescence as a negative control.
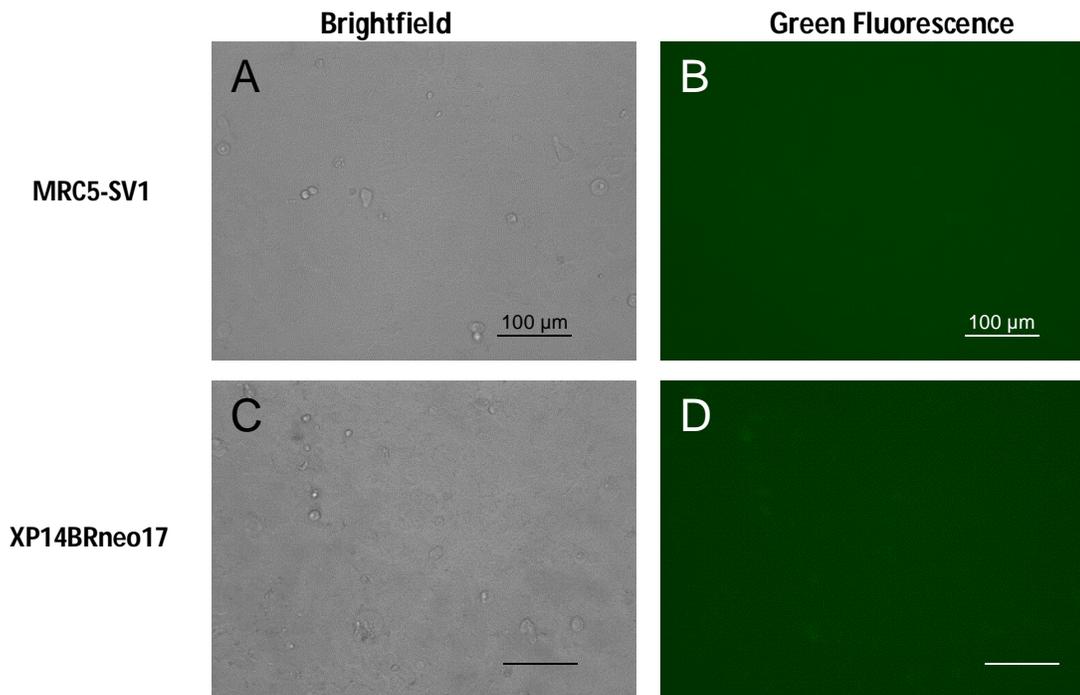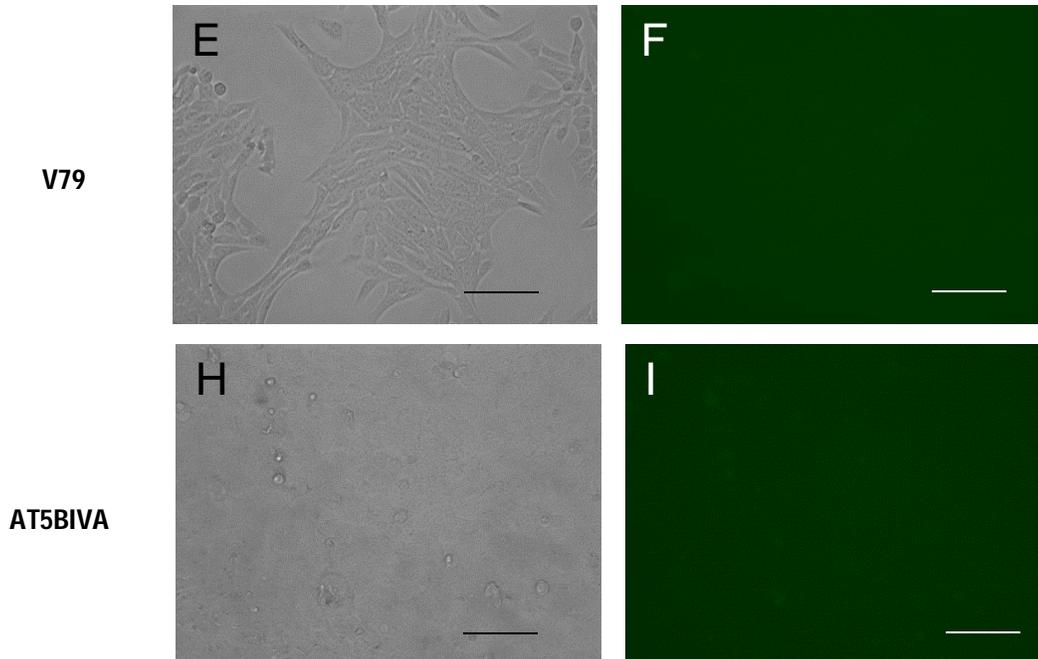
Figure 4.9 Cells exposed to green fluorescence inducing light prior to transduction. Cells are shows in brightfield microscopy (**A** + **C** + **E** + **H**) and 488 nm light (**B** + **D** + **F** + **I**). No visible fluorescence is seen in any of the cell lines: MRC5 (**A** + **B**), XP14 (**C** + **D**), V79 (**E** + **F**) and AT5 (**H** + **I**). These figures can act as negative controls. The photomicrographs were taken at 20x magnification with an EVOS FLoid Imaging System. Scale bars = 100 µm.

As seen in **figure 4.9**, prior to transduction with the GFP retroviral vectors, no visible green fluorescence is observed prior to transfection, thus these pictures serve as negative controls. Flow cytometry was used to quantify this lack of green fluorescence.

C: V79

D: AT5BIVA

Sample Statistics of Negative Control

Gate	Count	% Parent	X	Y	Median X	Median Y
All	80,000					
Cells	4,852	6.07 %	FSC-H	SSC-H	2,390,273	1,716,693
No Grn	4,852	100.00 %	FITC-H	FSC-H	5,670	2,390,273
Grn	0	0.00 %	FITC-H	FSC-H	0	0

Sample Statistics of AT5 Negative Control

Gate	Count	% Parent	X	Y	Median X	Median Y
All	15,000					
Cells	11,061	73.74 %	FSC-H	SSC-H	1,205,866	1,234,362
Green	139	1.26 %	FITC-H	FSC-H	549,333	1,527,950

Figure 4.10 Flow cytometry of sample cells prior to infection. All four cell lines are indicated next to their figure labels, (**A-D**). All cell lines indicate a very low amount of cells with the FITC-H signal sufficient to indicate green fluorescence. The green rectangles indicate where the cells would need to appear to count as having green fluorescence. Results were obtained using the NovoCyte Flow Cytometer and the data analysed using the NovoExpress software v1.2.5.

The results seen in **figure 4.10** corroborate those in **figure 4.9**. That there is indeed a low level of green fluorescence in the MRC5, XP14, V79 and AT5 line prior to transduction with *GFP*. Cell lines XP14 and V79 show 0% green fluorescence and the MRC5 line nearly the same at just 0.04% fluorescence. The AT5 cell line show some fluorescence, but comparable to that of the negative control sample seen in **figure 4.8 B** at simply 1.26%. These samples can act as negative controls for further experiments.

4.4.2 Transduction with *GFP* using Retroviral Vectors

4.4.2.1 First Transduction Attempts with Producer Line Titres

The first transductions were attempted in 2 batches on MRC5-SV1, XP14BRneo17 and V79 cell lines with consistent titres obtained from the SRH producer cell line as described in **section 2.2.6**.

MRC5 SV1

First Instance — Brightfield / Green Fluorescence

Repeat — Brightfield / Green Fluorescence

XP14BRneo17

| **First Instance** | | **Repeat** | |
| Brightfield | Green Fluorescence | Brightfield | Green Fluorescence |

V79

| **First Instance** | | **Repeat** | |
| Brightfield | Green Fluorescence | Brightfield | Green Fluorescence |
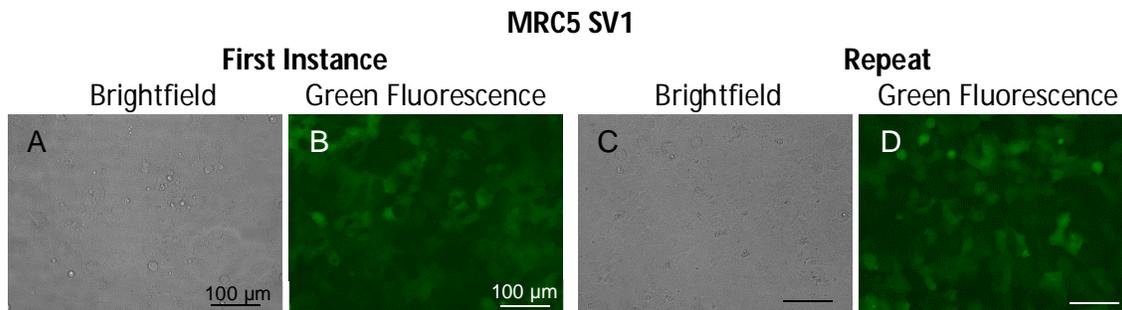
Figure 4.11 Cells exposed to green fluorescence inducing light following *GFP*-LV transduction. The cell lines were transduced in two batches with each brightfield image corresponding to the green fluorescence image to the right. The MRC5 (**B** - **D**), XP14 line (**F + H**) and the V79 (**I** - **L**) are indicated. Infections were done at consistent titres with the method described in **section 2.2.5**. Photomicrographs were taken at 20x magnification with an EVOS FLoid Imaging System. Scale bars = 100 μm.

The MRC5 and V79 cell lines appear to show some fluorescence as seen in **figure 4.11**. However, this fluorescence appears quite dull compared to those seen of the positive control cell line in **figure 4.7 B**. The XP14 cell line showed no fluorescence. Flow cytometry analysis was performed on the cells to quantitatively determine the amount of green fluorescence.

A: MRC5-SV1

Sample Statistics of SRH GFP

Gate	Count	% Parent	X	Y	Median X	Median Y
All	15,000					
Cells	11,194	74.63 %	FSC-H	SSC-H	2,448,606	1,036,851
No Grn	10,406	92.96 %	FITC-H	FSC-A	3,163	1,550,509
Grn	779	6.96 %	FITC-H	FSC-A	174,198	2,021,246

B: XP14BRneo17

Sample Statistics of SRH GFP

Gate	Count	% Parent	X	Y	Median X	Median Y
All	80,000					
Cells	4,686	5.86 %	FSC-H	SSC-H	2,411,169	1,831,866
No Grn	4,686	100.00 %	FITC-H	FSC-H	6,401	2,411,169
Grn	0	0.00 %	FITC-H	FSC-H	0	0

C: V79

Sample Statistics of SRH GFP

Gate	Count	% Parent	X	Y	Median X	Median Y
All	15,000					
└─ Cells	11,249	74.99 %	FSC-H	SSC-H	2,354,187	940,021
├─ No Grn	11,002	97.80 %	FITC-H	FSC-H	4,464	2,349,318
└─ Grn	249	2.21 %	FITC-H	FSC-H	143,390	2,566,401

Figure 4.12 Flow cytometry of cell samples following GFP transduction. A, **B** and **C** are flow cytometry results of the cell lines MRC-5SV1, XP14BRneo17 and V79 respectively following transduction with LV GFP produced by the SRH cell line. Results were obtained using the NovoCyte Flow Cytometer and the data analysed using the NovoExpress software v1.2.5. Student's t tests indicated significant difference between all corresponding results with a p value lower than 0.01.

As seen in **figure 4.12**, all three lines show very little GFP transduction, indicated with the cells appearing in the green rectangles. MRC5 showed 6.96% green fluorescence, XP14 showed 0% fluorescence and V79 showed 2.2% fluorescence.

The low amount of fluorescence in both photomicrography (**figure 4.11**) as well as flow cytometry (**figure 4.12**) indicates poor GFP transduction. Further experiments were required to determine the necessary optimisation to allow for the successful transduction of GFP.

4.4.2.2 VSV-G Envelope Plasmid Supplementation

In order to successfully transduct cells, the VSV-G viral envelope protein was considered to be utilised to allow for better transduction of cells. The envelope glycoproteins are known to form complexes with plasmid DNA and MLV retrovirus particles to improve transduction as well as improve fusion of the viral envelope with the host cell membrane, allowing for better insertion of viral DNA into the cell (Ci *et al.*, 2018). Plasmids containing the *VSVG* gene were added at a ratio of 10 plasmids for every host cell after the retroviral agent was applied to them. This was attempted on the MRC5-SV1, XP14BRneo17 and V79 cell lines.

MRC5 SV1

First Instance		**Repeat**	
Brightfield	Green Fluorescence	Brightfield	Green Fluorescence

Xp14BRneo17

First Instance		**Repeat**	
Brightfield	Green Fluorescence	Brightfield	Green Fluorescence

V79

First Instance		**Repeat**	
Brightfield	Green Fluorescence	Brightfield	Green Fluorescence
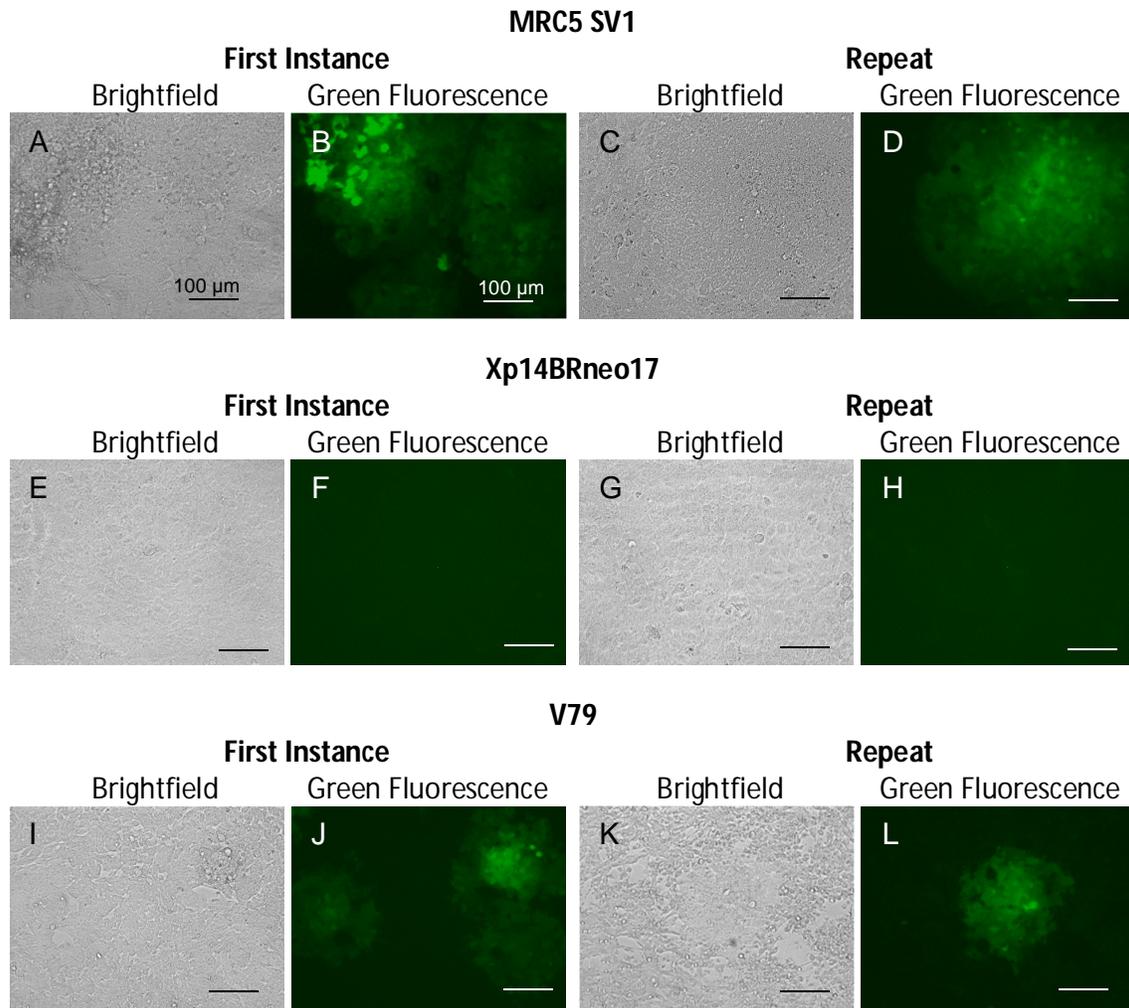
Figure 4.13. Cells exposed to green fluorescence inducing light following VSV-G supplemented *GFP*-LV transduction. The cell lines were transduced in two batches with each brightfield image corresponding to the green fluorescence image to the right. Each row of images corresponds to the cell line indicated above: **A-D** are MRC cells, **E-H** are XP14 cells and **I-L** are V79 cells. Infections were done with consistent titre with the method described in the **materials and methods section 2.2.5.** The photomicrographs were taken at 20x magnification with an EVOS FLoid Imaging System. Scale bars = 100 μm.

Fluorescence in **figure 4.13**, if at all present, was faint compared to the positive control of the SRH producer cell line as seen in **figure 4.7 B**. The XP14 cell line (**E – H**) showed no fluorescence comparable to the amount seen than the transduction attempted without VSV-G as seen in **figure 4.11** (**F + H**). The MRC cell line (**A - D**) showed the highest amount of green fluorescence in a large amount of visible cells, more than the attempt in **figure 4.11**. The V79 cell line showed a moderate amount of fluorescence as compared to the other two cell lines with slightly more fluorescence compared to **figure 4.11** (**J +**

L). Both MRC5 and V79 cell lines appear to have altered cell morphology following transduction as compared to **figure 4.11**.

Flow cytometry analysis was performed on the cells to determine quantitative results.
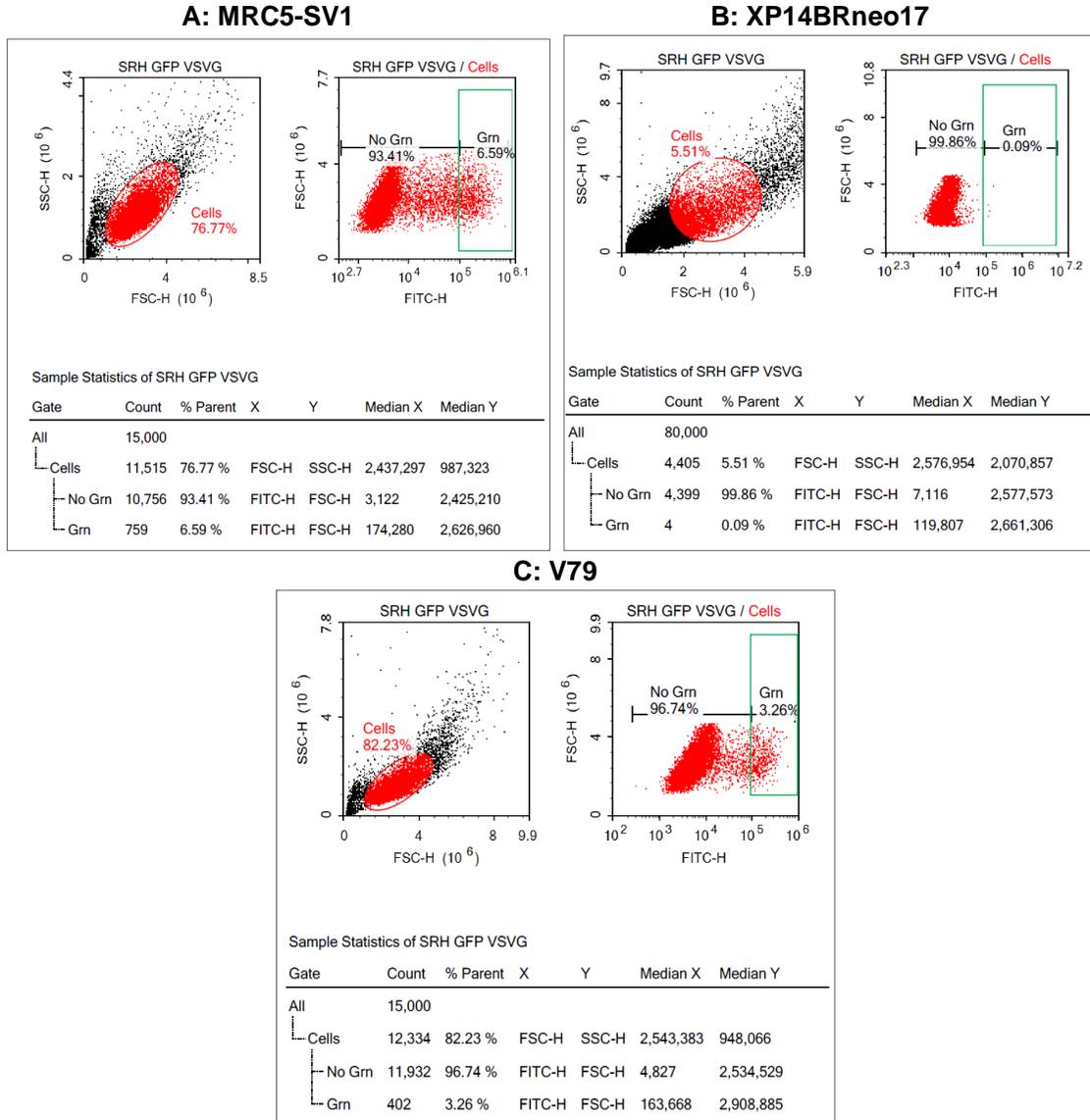
Figure 4.14 Flow cytometry of sample following VSV-G supplemented *GFP*-LV transduction. A, **B** and **C** are the cell lines MRC-5SV1, XP14BRneo17 and V79 respectively following transduction with LV GFP produced by the SRH cell line and incubated with VSV-G before transduction. The green rectangles indicate where the cells would need to appear to count as having green fluorescence. Results were obtained using the NovoCyte Flow Cytometer and the data analysed using the NovoExpresss software v1.2.5.

Results seen in **figure 4.14** corroborate those seen in figure **4.13**. MRC5 (**4.13 C**) the most fluorescence at 6.59%, a very small dip from the first experiment done without

94

VSV-G as seen in **figure 4.12 A** (Student's T test indicated no significant difference with a p value above 0.05). The remaining cell lines showed a little improvement. XP14 (**4.14 B**) showed 0.09% fluorescence up from 0% (**4.12 B**), and V79 (**4.14 B**) showing 3.26% fluorescence up from 2.21% (**4.12 C**). While **figure 4.13** indicated that the cells were expressing GFP, quantification of the fluorescence showed a low amount of fluorescence (Student's T test indicated significant difference with a p value less than 0.01). The V79 and MRC cell lines also showed altered morphology following infection with VSVG as seen in **figure 4.13**.

4.4.2.3 Retroviral Vector Aliquot Transductions

Due to apparently poor transduction even with VSVG supplementation of the consistent producer cell titres, *GFP* transductions were attempted with known titres of retroviral vectors at an MOI of 10 as described in **section 2.2.6**. The experiments were repeated with two separate batches of *GFP*-LV aliquots of known titres provided by Saqlain Suleiman (2018) and by Dr. Michael Themis (2018). These experiments were performed on the MRC5-SV1 and AT5BIVA cell lines. The cells were first imaged prior to transduction.
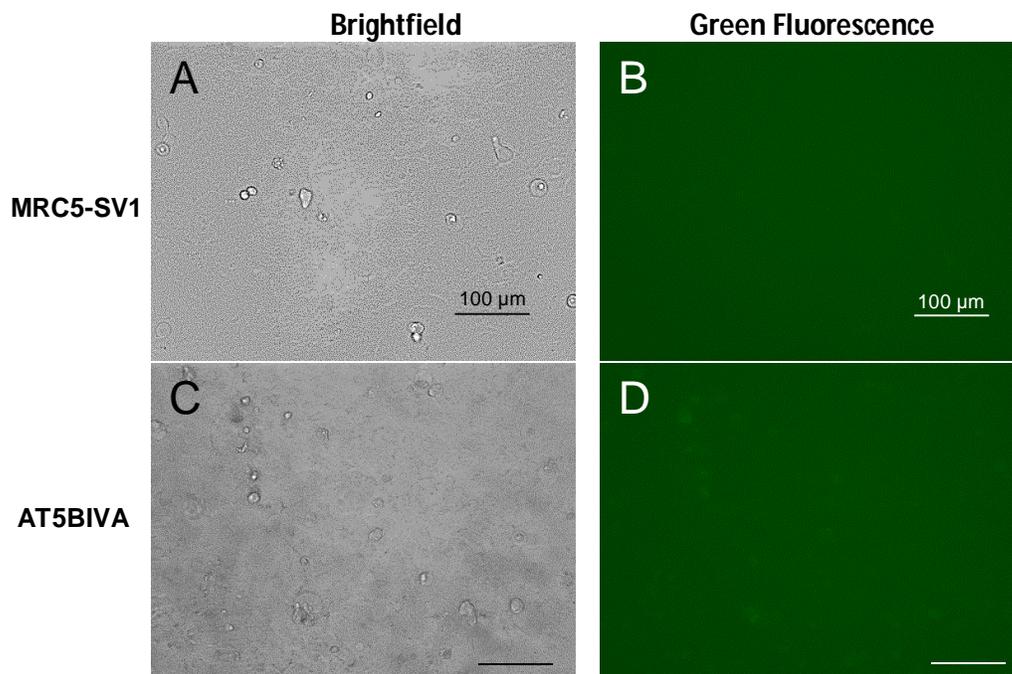
Figure 4.15 Cells exposed to green fluorescence inducing light following transduction. The images to the left (**A+C**) show brightfield microscopy while the ones on the right (**B+D**) show cells under ultraviolet light. The upper row of images (**A+B**) are the MRC line while the bottom (**B+D**) are the AT5 line. Infections were done at an MOI of 10 with the method described in the **materials and methods**

Figure 4.15 shows negative control results of the MRC5 and AT5 lines prior to infection and appear to show no fluorescence. The cells were ran through a flow cytometer to quantify any green fluorescence.
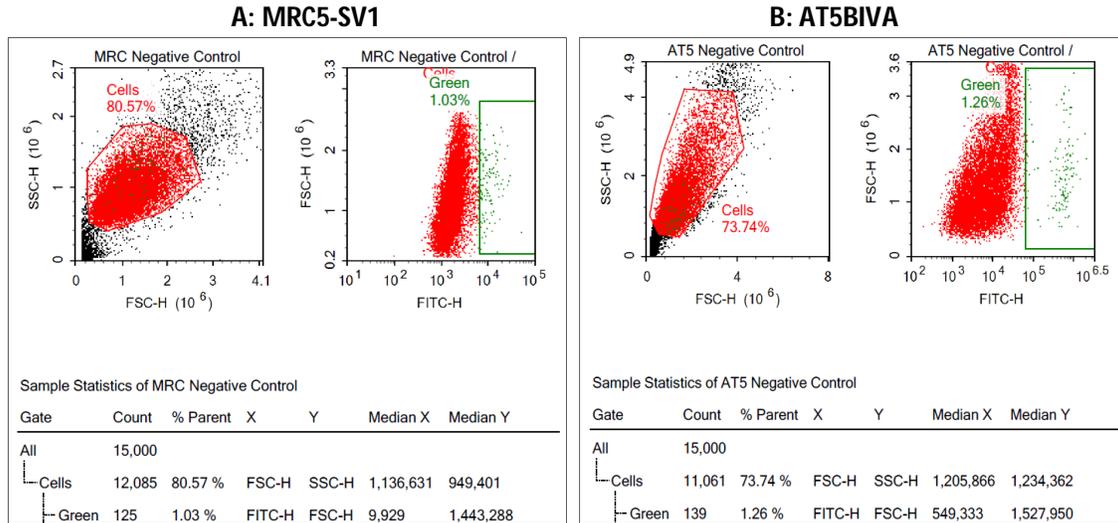
A: MRC5-SV1   **B: AT5BIVA**

4.16 Flow cytometry of MRC-5SV1 and AT5BIVA cells prior to infection. The FITC-H signal sufficient to indicate green fluorescence are highlighted with green rectangles. **A** shows the MRC5 cell line results while **B** shows the AT5 cell line results. Infections were done at an MOI of 10 with the method described in the **section 2.2.5**. Results were obtained using the NovoCyte Flow Cytometer and the data analysed using the NovoExpress software v1.2.5

As seen in **figure 4.16**, the MRC5 line (**A**) showed 1.03% of cells with green fluorescence, while the AT5 cells (**B**) showed 1.26% of cells with green fluorescence. The cells were then transducted in two batches with GFP LV aliquots of known titres provided by Dr. Saqlain Suleman (2018) at an MOI of 10.

	Brightfield	**Green Fluorescence**
MRC5-SV1	A	B
AT5BIVA	C	D
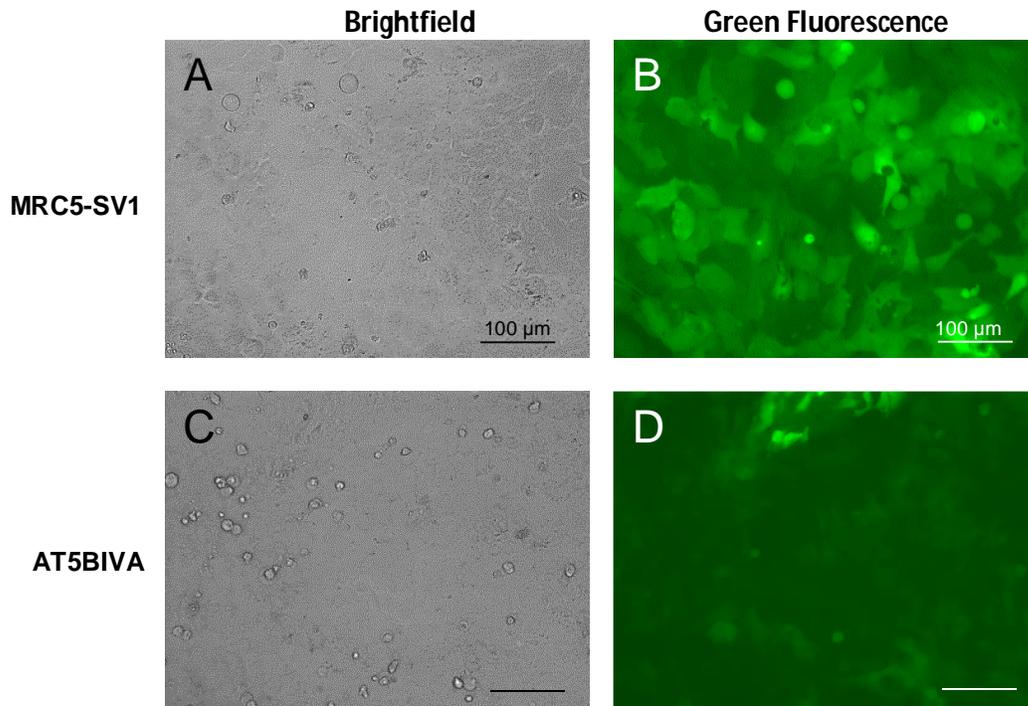
4.17 Batch 1 of Cells following GFP LV aliquot transduction. Batch 1 of MRC5 and AT5 cell lines transducted with *GFP*-LV and visualised in green fluorescence inducing light. **A** and **C** show brightfield microscopy of the cells. **B** and **D** show the cells under ultraviolet light. **A** and **B** show the MRC cell line while **C** and **D** show the AT5 cell line. The cells were transduced with *GFP*-LV aliquots provided by Dr. Saqlain Suleman (2018). Infections were done at an MOI of 10 with the method described in **section 2.2.5**. The photomicrographs were taken at 20x magnification with an EVOS FLoid Imaging System. Scale bars = 100 μm.

There is more green fluorescence seen in these figures as compared to the control samples which were not transducted with *GFP*-LV as seen in **figure 4.15** indicating successful transduction. The green fluorescence seen appears to be stronger than in **section 4.4.2.2** done with consistent titres obtained from producer cells, indicating the viral aliquots to be a better for transduction. There appears to be more fluorescence seen in the MRC5 cell line (**figure 4.17 B**) compared to the AT5 line (**figure 4.17 B**).

The cells were run through a flow cytometer to quantify the green fluorescence. This was done in two sets as a large amount of cells were grown.
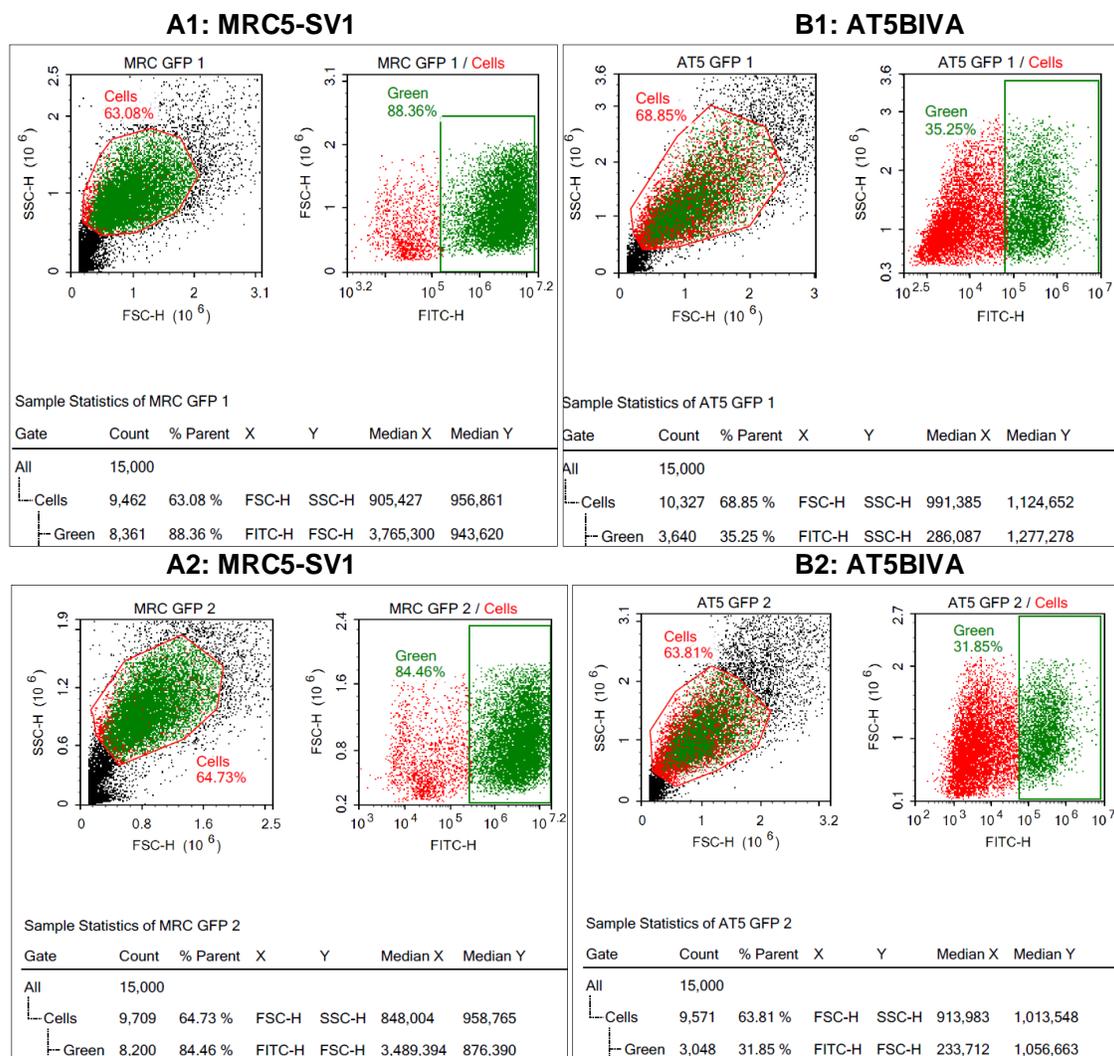
Figure 4.18 Flow Cytometry of batch 1 of MRC5 and AT5 cells following GFP LV aliquot transduction. A1 and **A2** are flow cytometry results of the MRC-5SV1 cells while **B1** and **B2** are flow cytometry results of the AT5BIVA cells. The green rectangles indicate the area in which cells must be to count as having green fluorescence. Infections were done at an MOI of 10 with the method described in the **materials and methods section 2.2.5**. Results were obtained using the NovoCyte Flow Cytometer and the data analysed using the NovoExpress software v1.2.5.

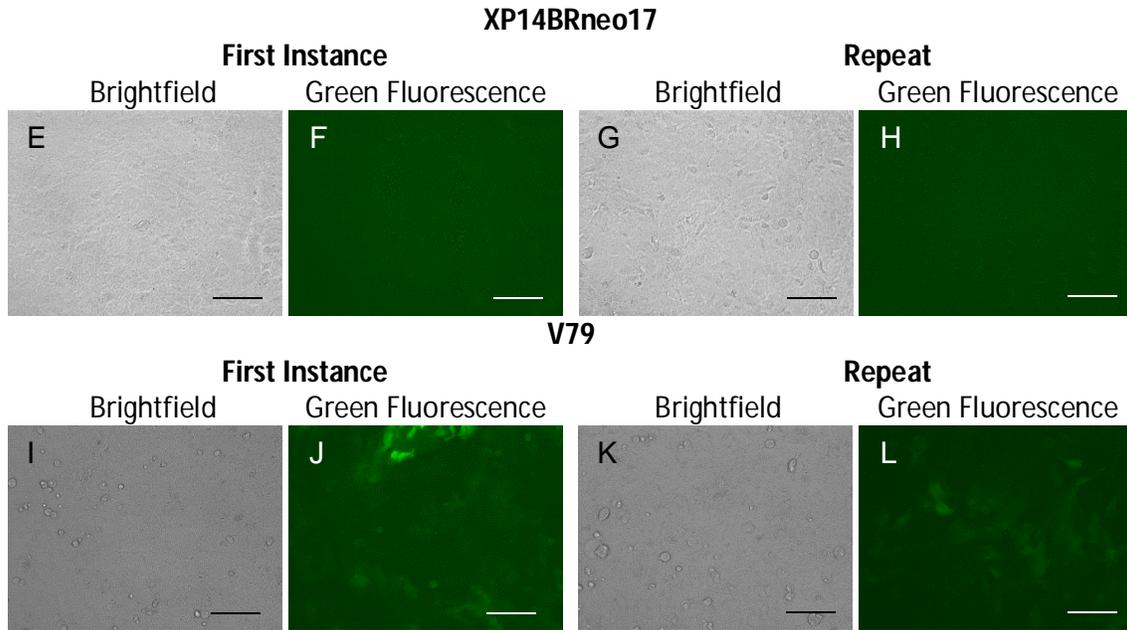The results in **figure 4.18** corroborate the results seen in **figure 4.17**. Both MRC5 and AT5 cell lines indicate a marked increase in the cells indicating fluorescence from the control samples seen in **figure 4.16**. MRC cells shown in **4.18 A1** and **A2** respectively show 88.03% and 84.46% of cells with green fluorescence. AT5 cells shown in **4.18 B1** and **B2** respectively show 35.25% and 31.85% cells with green fluorescence. The MRC5 cells have shown better transduction than the AT5BIVA cells. The experiment was repeated on the MRC5-SV1 and AT5BIVA cells with a separate batch of *GFP*-LV aliquots from a different source to confirm the results.
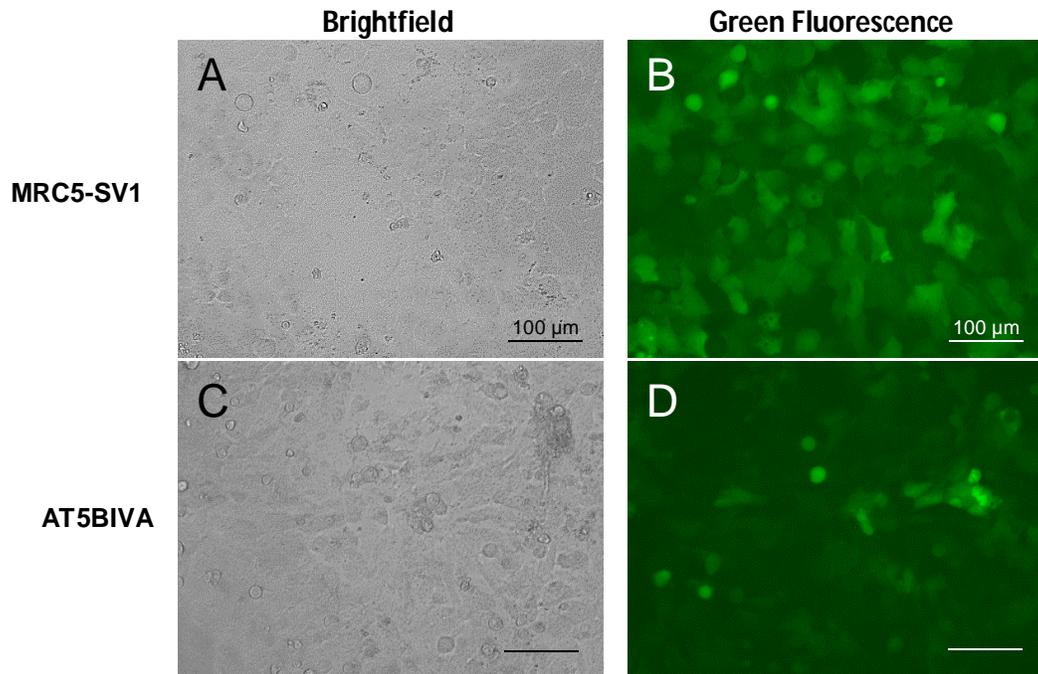
Figure 4.19 Batch 2 of Cells following GFP LV aliquot transduction. Batch 2 of MRC5 and AT5 cell lines transducted with *GFP*-LV and visualised in green fluorescence inducing light. **A** and **C** show brightfield microscopy of the cells. **B** and **D** show the cells under ultraviolet light. **A** and **B** show the MRC cell line while **C** and **D** show the AT5 cell line. The cells were transducted with *GFP*-LV aliquots provided by Prof. Michael Themis (2018). Infections were done at an MOI of 10 with the method described in the **materials and methods section 2.2.5**. The photomicrographs were taken at 20x magnification with an EVOS FLoid Imaging System. Scale bars = 100 μm.

There is more green fluorescence seen in **figure 4.19** (**B** and **D**) as compared to the negative control samples which were not transducted with *GFP*-LV as seen in **figure 4.15** indicating successful transduction. The AT5 cells (**figure 4.19 D**) appear brighter than those done in the first instance of the experiment seen in **figure 4.17 D**. There is more fluorescence seen in the MRC5 cell line, similar to **figure 4.17 B** indicating better transduction.

The cells were run through a flow cytometer in 2 batches to quantify the green fluorescence.
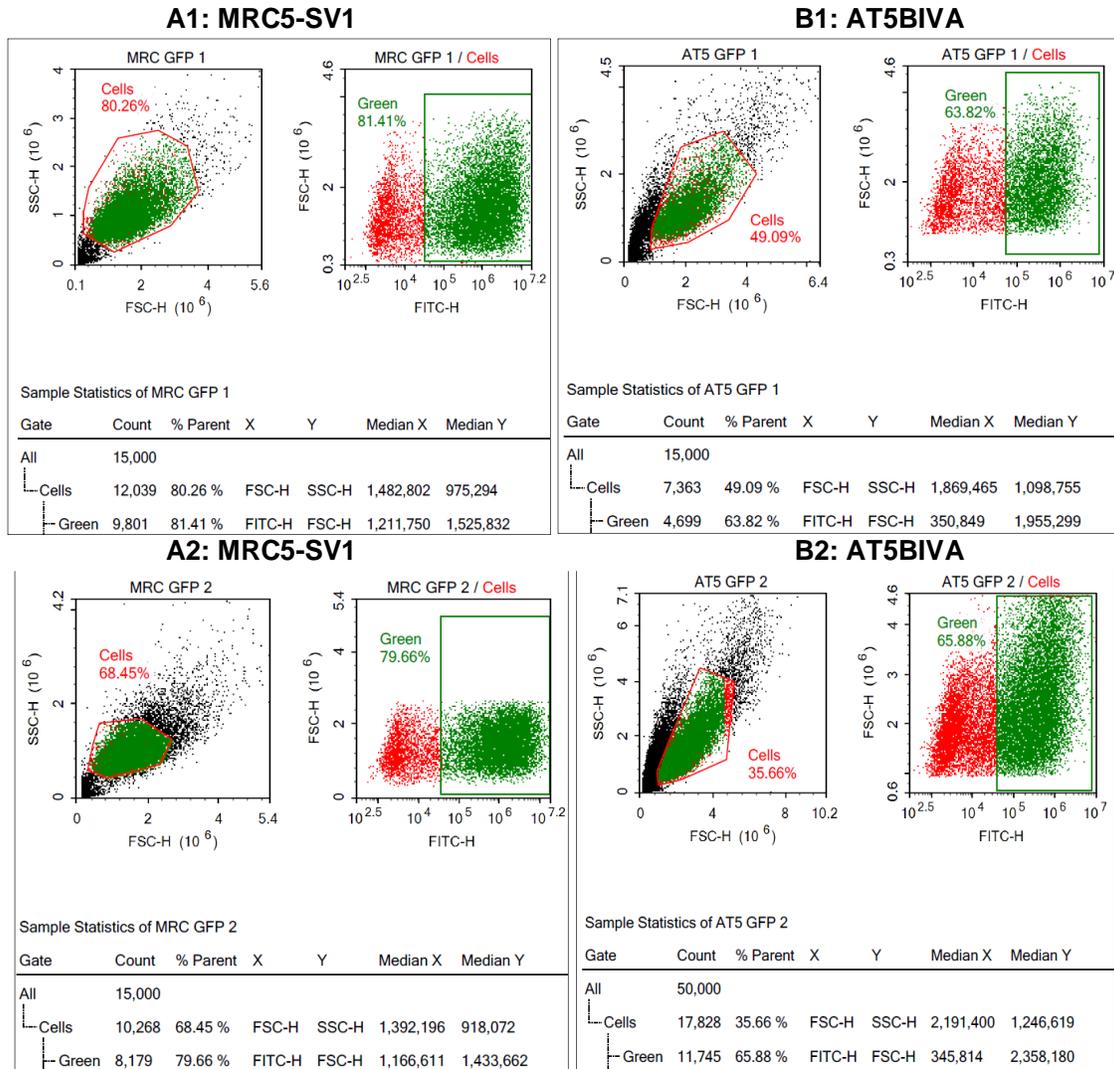
Figure 4.20 Flow Cytometry of batch 2 of MRC5 and AT5 cells following *GFP* LV aliquot transduction. A1 and **A2** are flow cytometry results of the MRC-5SV1 cells while **B1** and **B2** are flow cytometry results of the AT5BIVA cells. The green rectangles indicate the area in which cells must be to count as having green fluorescence. Results were obtained using the NovoCyte Flow Cytometer and the data analysed using the NovoExpress software v1.2.5.

Similar to the results seen in **figure 4.18**, in the results shown in **figure 4.20**, both cell lines indicate a marked increase in cells indicating fluorescence from the control samples seen in **figure 4.18**. The MRC line, **A1** and **A2**, respectively show 81.41% and 79.66% of cells with green fluorescence. The AT5 line, **B1** and **B2**, respectively show 63.82 and 65.88% cells with green fluorescence. The AT5BIVA cells had an increase in fluorescence as compared to that seen in **figure 4.18 B1** and **B2**; an increase from 35.25% and 31.85% to 63.82 and 65.88% indicating better transduction. The AT5 cells still show a reduced amount of fluorescence compared to the MRC5 cells.

Table 4.1 below shows a compilation of the results of the two batches of results seen in **figure 4.18** to **4.20**.

> **Table 4.1 Compilation of Flow Cytometry Results seen in figure 4.18 and 4.20.** A combined % of green fluorescent cells was obtained because the transductions on both repeats were performed at the same time. Student's T tests were used to determine differences between samples and are indicated below the figure.

Flow Cytometry Results Compiled		MRC5-SV1		AT5BIVA	
		Repeat 1	**Repeat 2**	**Repeat 1**	**Repeat 2**
Control	**Total Cells**	12085		11061	
	Green Cells	125		139	
	% Green Cells	**1.03%**		**1.26%**	
GFP-LV 1	**Total Cells**	9462	9709	10327	9571
	Green Cells	8361	8200	3640	3048
	% Green Cells	88.36%	81.41%	35.25%	63.82%
	Combined %	**84.47%**		**47.14%**	
GFP-LV 2	**Total Cells**	12039	10268	7363	17828
	Green Cells	9801	8179	4699	11745
	% Green Cells	84.46%	79.66%	31.85%	65.88%
	Combined %	**81.99%**		**53.99%**	

As seen in **table 4.1**, the MRC5 repeats were very close to each other (Student's T test indicated no significant difference with a p value above 0.05), however there was in the ATBIVA cells indicating better transduction in the latter experiment (Student's T test indicated significant difference with a p value below 0.01). The AT5 cells also show poorer transduction than the MRC5 line. The AT5 cells have an impaired DNA DSB repair mechanism, while MRC5 has an intact one, thus indicating the importance of DSB repair for retroviral vector integration. However, both cell lines showed significant increase in fluorescence as compared to the control samples shown in **figure 4.16** (Student's T test indicated significant difference with a p value below 0.01).

Following these experiments, it was determined that the *GFP*-LV aliquots were adequately transducting target cells.

4.5 GFP Vector Infection of HAT Selected Cells

As seen in **figures 4.4**, **4.5** and **4.6** under **section 4.3**, following HAT selection, the MRC5-SV1 and V79 were the two cell lines which adequately produced the required amount of cells necessary for the second selection of the HPRT assay. Thus, a simple GFP profile was produced with the two cell lines after going through HAT selection to show that they can be transducted with the *GFP*-LV.
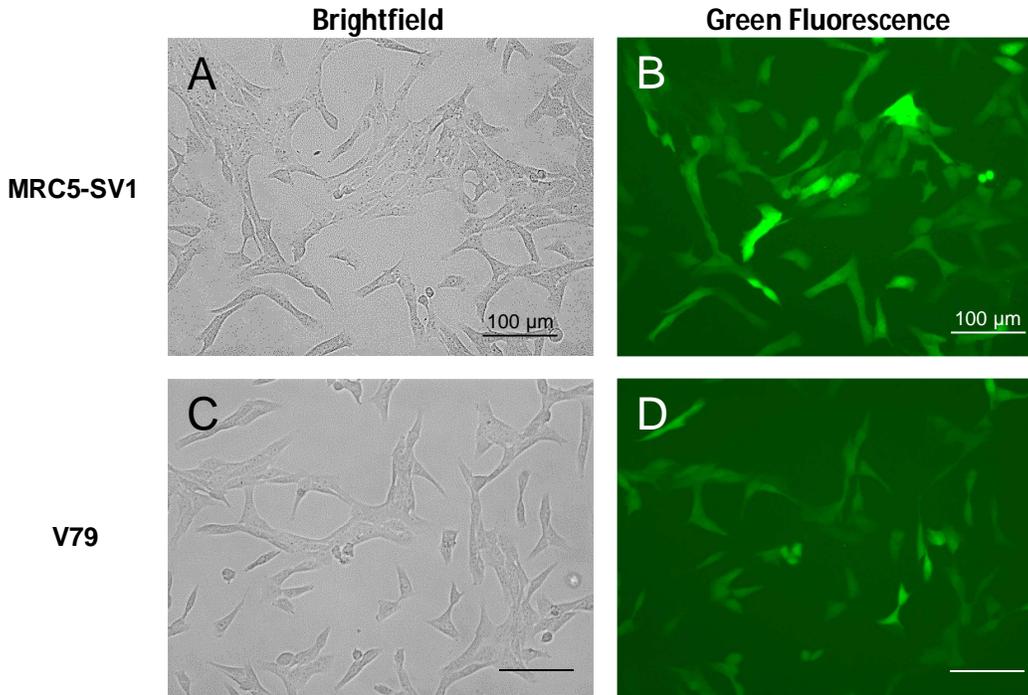
Figure 4.21 HAT selected MRC5 and V79 cells following GFP LV aliquot transduction.
Fluorescence is seen in both cell lines following transduction. Infections were done at an MOI of 10 with the method described in the **materials and methods section 2.2.5**. The photomicrographs were taken at 20x magnification with an EVOS FLoid Imaging System. Scale bars = 100 μm.

The cells were run through a flow cytometer to quantify the green fluorescence.
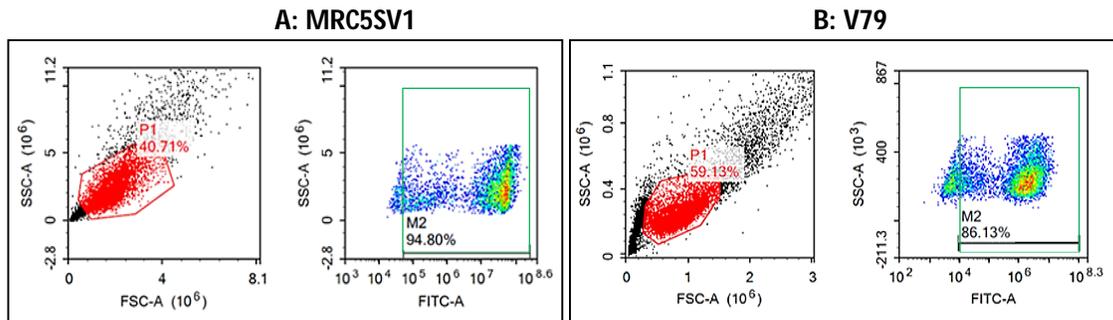
Figure 4.22 Flow Cytometry results of HAT selected MRC5 and V79 cells following GFP LV aliquot transduction. A) and **B**) show the MRC5 and V79 results respectively. P1 indicates the sample being gated for single cells as opposed to cell debris or multiple cells clumped together. The green rectangle, M2, indicates the percentage of cells within the P1 gates which display green fluorescence. Results were obtained using the NovoCyte Flow Cytometer and the data analysed using the NovoExpress software v1.2.5.

MRC5 cells (**figure 4.22 A**) showed 94.8% cells which were fluorescent. V79 cells (**figure 4.22 B**) showed 86.13% cells which were fluorescent. As a marked increase in fluorescence was seen in the majority of cells from both MRC5-SV1 and V79 cell lines

compared to the negative control cells not transducted with GFP (**figure 4.14 A** and **C**), successful GFP transduction can be inferred.

4.6 8AZ and 6TG Selection

The MRC5-SV1 cells and V79 cells selected in HAT as shown in **figures 15.1** and **15.3** were each separated into eight colonies. Three of four pairs of each cell line were transducted with *GFP*-LV, gag-pol plasmids to encode for the integrase enzyme, and empty LV vectors. The fourth pair was not transducted with any material, serving as a negative control. Following transduction, each colony from each pair was grown in 8-azaguanine (8AG) or 6-thioguanine (6TG) for a period of 20 days as described in the **materials and methods section 2.2.13**. Following this, the number of colonies were counted, a higher number of colonies indicating a higher genotoxicity of the transduction vector being tested.
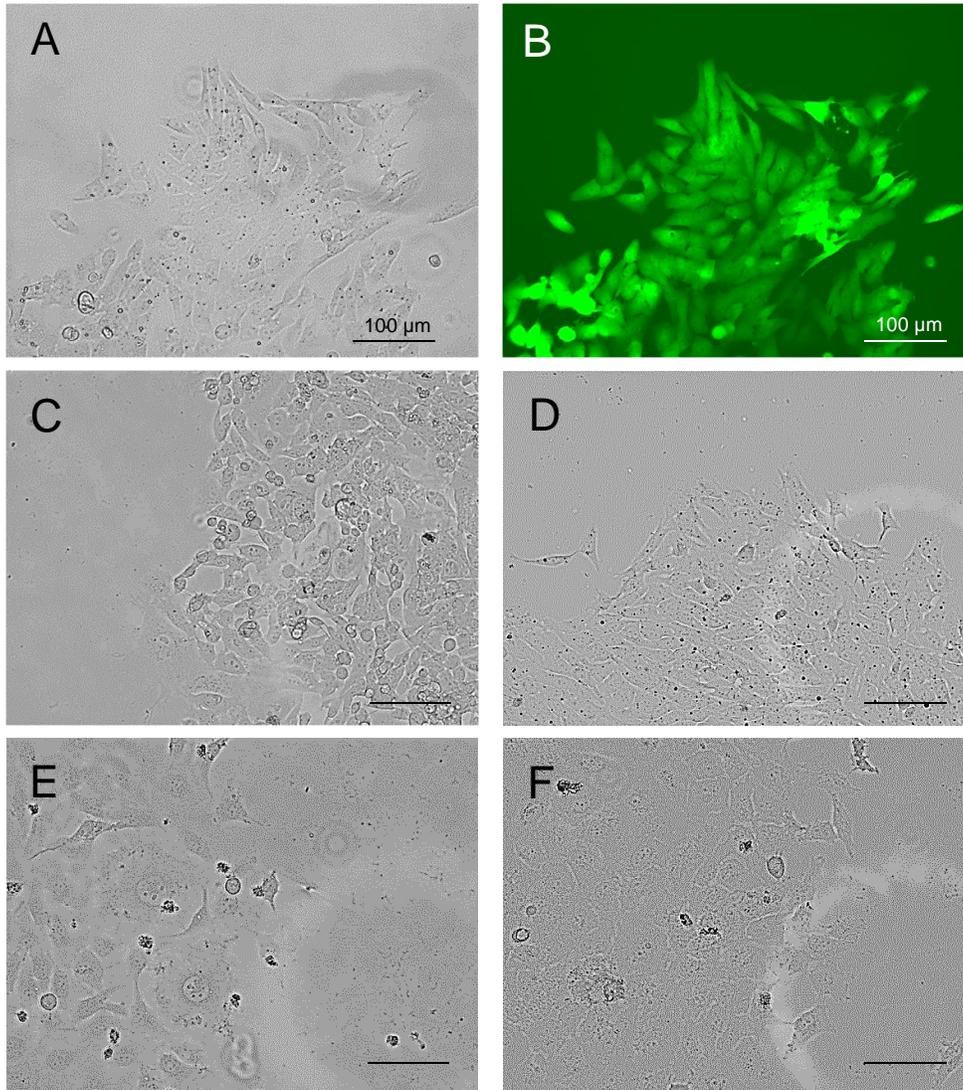
Figure 4.23 Examples of V79 cells selected in HAT medium growing in 8AZ and 6TG. Images of V79 cells at various points during the secondary selection of the *HPRT* assay. **B** was taken under ultraviolet light while the others are filmed under brightfield light. Photomicrographs **A** and **B** are examples of the same group of cells which have been transduced by GFP and growing in 8AZ medium. **C** shows cells with integrase protein obtained from *gag pol* plasmid expression applied in 8AZ. **D** shows empty vector transducted cells in 8AZ. **E** is an example of GFP transducted cells in 6TG secondary medium and **F** is an example of integrase from *gag pol* expression in 6TG selection medium. The cells were grown for 20 days. The selection was done as described in **section 2.2.13**. The photomicrographs were taken at 20x magnification with an EVOS FLoid Imaging System. Scale bars = 100 µm.

Green fluorescence is clearly seen in **figure 4.23 B**, indicating successful *GFP* transduction. The cells are growing normally in all the photomicrographs.

Table 4.2 Results of MRC5 and V79 cells selection in 8AG and 6TG. The table shows the number of colonies formed after secondary selection of cells in 8AG or 6TG. The cells would have undergone HAT selection and viral transduction prior to this step. The more the colonies that grow after a transduction, the more genotoxic they as the *HPRT* gene has been rendered non-functional, preventing the cells from utilising the salvage pathway of nucleotide synthesis with the toxic 8AG and 6TG.

Second *HPRT* Assay Selection Step	Cell type	V79	MRC5
Average number of Colonies grown in 6TG selection	Control	1.5	0
	GFP Virus	6	6
	Empty vector	1	1
	Gag – Pol/Integrase	13.5	8
Average number of Colonies grown in 8AZ Selection	Control	0	0
	GFP Virus	11	11
	Empty Vector	0.5	1
	Gag – Pol/Integrase	15.5	13

Table 4.2 shows the empty vectors indicating little genotoxicity as compared to the controls. The GFP had the second highest amount of genotoxicity and the integrase protein displayed the highest amount of genotoxicity. Graphic representation of the results are shown below in **figure 4.25**.
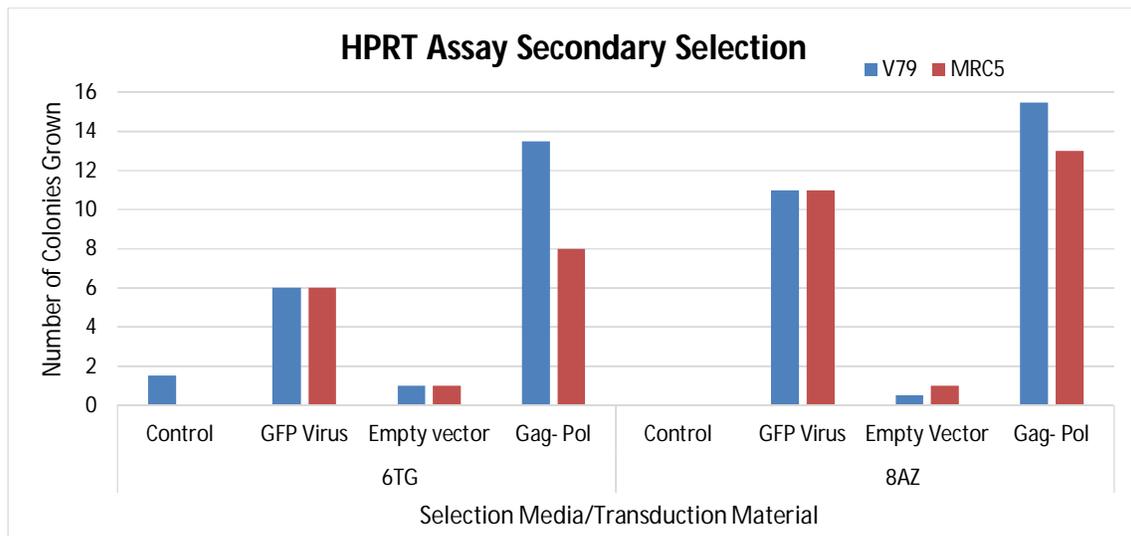
Figure 4.24 Graph showing results of MRC5 and V79 selection in 8AG and 6TG. A graph made from the data seen in **table 4.2**. The more the colonies that grow after a transduction, the more genotoxic they as the HPRT gene has been rendered non-functional, preventing the cells from utilising the salvage pathway of nucleotide synthesis with the toxic 8AG and 6TG.

4.7 Discussion

4.7.1 General Results

To determine the genotoxicity caused by viral vectors with and without genomes, the HPRT assay was used. The *HPRT* gene is present on the X chromosomes, only a single copy being present in male cells, thus providing a useful single gene knockout assay for genotoxicity.

HPRT is used in the salvage pathway of DNA synthesis. HAT medium was used to select for cells that only have functional *HPRT* by first blocking the *de novo* pathway of DNA synthesis with the drug with aminopterin. *HPRT* uses the salvage pathway to process hypoxanthine and thymidine from the HAT medium to allow the cell to continue synthesising nucleotides. This makes cells without functional *HPRT* die out, thus, following HAT selection, a population of cells without background *HPRT* mutations can be obtained (See **figure 4.1)**. These cells with no to very little background *HPRT* mutations were then infected with the viral vectors with and without genome to later determine the genotoxic effect of these infections.

Following infections, the cells undergo a second selection process with the toxic purine analogues 8AG and 6TG. Even with the *de novo* pathway is still active, the cells will still use the salvage pathway. Cells with functional *HPRT* will incorporate the toxic compounds into their DNA and cause cell death, and cells with a mutated *HPRT* will be able to survive. Thus any colonies formed following this secondary selection will come from cells which had genotoxicity knock out their *HPRT* gene; a count of these colonies will allow a quantifiable result to indicate the level of genotoxicity of the viruses with and without genomes used in between selections (See **figure 4.2**) (OECD, 2016; Themis *et al.*, 2003; Grosovski *et al.*, 1993; King *et al.*, 1985; Bradley *et al.*, 1981).

Before the secondary selection in 8AG or 6TG, the infection of the host cells by *GFP*-vector was first optimised. First, consistent tires (lacking VSVG) from producer cells were attempted and showed little integration. This integration was attempted to be boosted by adding the VSVG envelope proteins, which made little difference. Following this, aliquots of viruses with known titres already containing the VSVG proteins in their envelopes protocol were utilised at an MOI of 10, following which GFP transduction was seen. Thus it was decided to use viruses already formed with VSVG envelopes rather than adding VSVG to infections done with non-VSVG viruses of unknown titres.

High amounts of genotoxicity was seen in cells infected by the GFP virus as compared to the control samples and a higher amount seen with gag-pol transfected cells. The Empty vectors genotoxicity was comparable to the control samples which were not infected. Both cells lines were NHEJ competent and had comparable amounts of genotoxicity in the control samples and when infected by the GFP virus. The V79 cells showed a comparatively higher amounts of genotoxicity compared to the MRC5 cell line when transfected with the gag-pol plasmids. 8AZ showed a larger number of colonies formed compared to 6TG, indicating that 8AZ is a more sensitive selection agent for the HPRT assay. These results are presented in **table 4.2** and **figure 4.23**.

In regards to "complete" vectors, the results align with those seen in **chapter 3** where DSB formation was high in infections done with complete vectors. This shows that DSB formation could be linked to genotoxicity. However, empty vectors showed very low genotoxicity, comparable to the negative control samples where no transduction was performed. This may mean that integration itself is a greater cause of genotoxicity as "complete" viral vector sequences may interrupt existing genes in host cells, meanwhile, "empty" vectors simply cause DSBs which are simply repaired since no genetic sequence is present to interrupt host gene function.

However, this is disputed by the *gag-pol* transfection causing the greatest amount of genotoxicity. In this case, the integrase protein coded by the *pol* gene would be causing damage; much like the empty vectors, there should be no viral vector genome to interrupt host cell sequences. Yet the integrase causes the highest genotoxicity; this may be due to the high number of integrase enzymes produces, leading to many more DSBs formed compared to complete or empty vectors which might generally be regulating integrase numbers. As such, due to a very high number of DSBs formed, the host cell needs to rely on the error prone NHEJ mechanism to repair the damage, which may cause disruptions in host cell sequences.

While prior publications have tested the genotoxic effects of lentiviral vectors on genotoxicity (Everson *et al.*, 2016; Cesana *et al.*, 2014), there is not much research on the genotoxic effects of empty vectors. These empty vectors may be formed as part of retroviral vector batches.

Prior research has also shown that lentiviral vectors can cause genotoxicity (Montini *et al*, 2009; Poletti and Mavilio, 2018), however, research has not been carried on the direct genotoxic effect of unregulated *gag-pol* expression on host cells. This also means that care must be taken in altering the regulatory parts of the retroviral genome when it is

altered for the purposes of retroviral gene therapy; it could be possible certain alterations made to the genome could cause unregulated expression of the viral integrase, leading to both DSB damage as well as genotoxic damage to gene therapy patients.

4.7.2 Further Work

Due to time constraints, a male cell line that was NHEJ incompetent which could survive HAT selection was unable to be utilised. To obtain data observing the effects of infection on the *HPRT* gene in NHEJ incompetent cell lines, further optimisation of HAT selection could be attempted to determine if ATBIVA could survive. Otherwise, more NHEJ incompetent male cell lines could be procured to see if any could survive HAT selection, or the NHEJ process could be targeted and impaired in a male cell line known to survive HAT selection, such as MRC5SV1 or V79, to form a new cell line which could then undergo the HPRT assay.

The γH2AX DSB repair profiles of *gag-pol* transfections could be further investigated and compared to complete viral vectors and empty vectors to determine if the *gag-pol* transfection can cause a higher amount of DSBs, thus supporting the theory that DSBs are the cause of genotoxicity when it comes to *gag-pol* transfections, and possibly insertion the cause when it comes to complete viral vectors. Analysis on retroviral vector batches could be done to determine if empty vectors are also packaged.

Methodologies to improve on VSVG supplementation to retroviral vector transduction could be done to improve the efficiency of retroviral vectors without the need for the vector itself to be packaged with the VSVG protein in its envelope.

Chapter 5 - Investigation of Human Endogenous Retroviruses (HERVs) in Retroviral Vectors

5.1 Background and Aims

Retroviral and lentiviral gene therapy vectors can cause DNA damage during the integration of genetic material into host DNA. **Chapter 3** of this thesis has demonstrated that DNA damage, in the form of DSBs, can be caused by retroviral vectors carrying a genome, and even by "empty" vectors which are lacking a genome. It was seen that this can be particularly harmful to individuals with impaired NHEJ mechanisms and the DSB repair is impaired. **Chapter 4** demonstrated that, while vectors with genomes are capable of causing genotoxicity at a level higher than that seen of control cells, empty vectors cause genotoxicity comparable to control cells which have not been infected with virus.

In summary: while vectors do not need a genome to cause DNA damage via DSB formation, a genome is likely necessary for gene knockout genotoxicity. Thus, if an empty vector could procure a viral genome, this new recombinant vector could cause genotoxicity.

One such source of viral DNA that empty vectors could procure are endogenous retroviruses (ERVs). This thesis focuses on ERVs which are found in humans and are known as human ERVs or HERVs. HERVs are thought to be partial to whole sequences of viruses which had infected humans ancestrally over millions of years; and over time have evolutionarily lost their infectious nature due to mutations and internal recombination (Steinhuber *et* al., 1995; Bushman *et al.*, 2005; Feschotte and Gilbert, 2012; Stoye, 2012; Dewannieux and Heidmann, 2013; Contreras-Galindo *et al.*, 2015; Wildschutte *et* al., 2016; Gifford *et al.* 2018; Khalfallah and Genge, 2021)

HERV-K113 is a HERV which is present in the human genome at 9,472 base pairs (bp) long (Turner *et al.*, 2001). It has been seen with its entire coding capacity including the *gag*, *pol* and *env* genes along with the 5' and 3' LTR regions (van der Kuyl, 2012; Beimford *et al.*, 2008). The fact that HERV-K113 has been seen to exist with its entire coding capacity indicates that it could potentially become involved in the retroviral vector integration process and cause genotoxicity if incorporated into empty retroviral vectors.

5.1.1 Aims and Objectives

- To show that HERV-K113 can be detected by PCR in human cells

- To determine whether HERV-K113 is packaged into virus particles and can be integrated into non-human cells using the PCR.

In order to accomplish these aims, a polymerase chain reaction (PCR) assay was established to indicate the presence of these HERVs. This PCR could be used to determine the presence of HERVs in non-human target cells which normally do not have HERVs following transduction with empty viral vectors produced in human cells.

Five primer pairs were designed to amplify a part of the provirus to determine presence within the human genome. All five amplify a sequence around 7,500 bp in length between the LTRs as seen in **table 5.1**. Three more primer pairs were designed to capture parts of the LTR as well and range around the 8,500 bp range. These primer pair product sizes are indicated in **table 5.1**. The primer sequences can be seen in the **materials and methods section** in **table 2.1**.

5.2 HERV PCR

The primer pairs were analysed for dimerisation to ensure they would not interact with each other to generate a band. A negative control was also tested containing the primers and all they need to function except for the target DNA sequence. This would demonstrate the lack of action of the primer if no target sequence is present. The primers were first tested on human DNA obtained from 293T cells.

Primers were designed to amplify a near 7,500 bp sequence of HERV K113 in between the LTR regions. Primers were also designed to incorporate around 8,500 bp sequence of the HERV which included part of the LTR regions. The sequence of the primers are indicated in the materials and methods section **2.1.8**. **Table 5.1** below indicates the sizes of the product expected from the PCR of each primer pair designed.

Table 5.1 Expected product sizes of each primer pair. Primers pairs were designed using the Primer 3 software by ELIXIR and analysed for dimerization using the Multiple Primer Analyzer software from Thermo Fisher Scientific.

HERV Primer Pair	Expected Product Size (BP)
HERV 1 / H1	7,534
HERV 2 / H2	7,526
HERV 3 / H3	7,513
HERV 4 / H4	7,500
HERV 5 / H5	7,456
LTR 1 / HL1	8,514
LTR 2 / HL2	8,765
LTR 3 / HL3	8,612

Figure 5.1 below shows the relative sizes of PCR product each primer pair was expected to produce. A detailed analysis of primer binding sites is show in **appendix 4**.
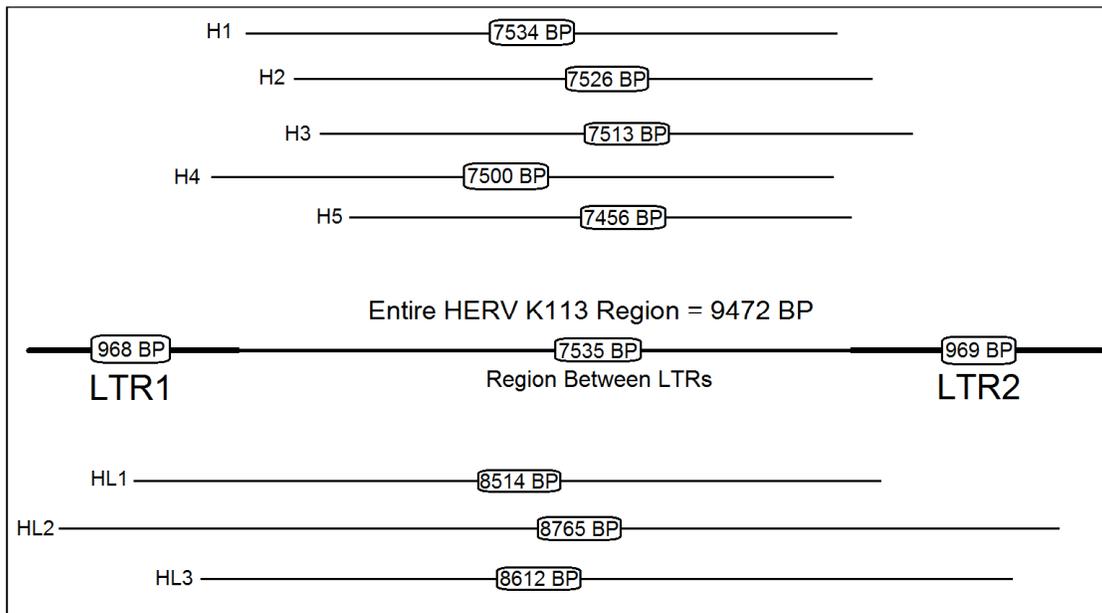
Figure 5.1 Diagram showing expected PCR products. Expected primer PCR products relative to primer binding sites. The relative size of the HERV K113 sequence is shown in the middle of the figure. The length the PCR products span relates to the sequence being amplified on the HERV sequence. Primer products above labelled **H1-H5** are amplifying regions between the LTRs, while those below labelled **HL1-HL3** are amplifying sequences including portions of the LTR regions. The diagram was produced using Microsoft Paint.

The PCRs were run as described in **section 2.2.16** unless stated otherwise. The PCRs were run on DNA extracted from the cell lines stated using the method described in **section 2.2.14**. Gel electrophoresis was used to show the results of the PCRs as described in **section 2.2.15**.

5.2.1 PCR Products of the HERV Primers
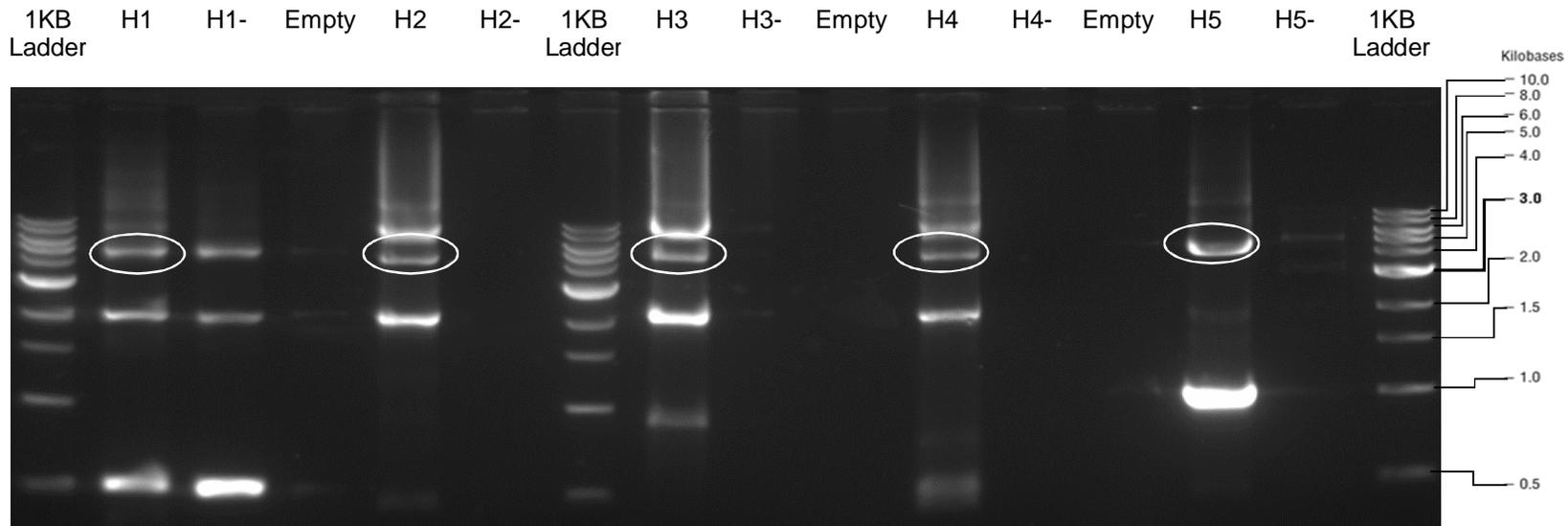
Figure 5.2 Gel Image of PCR-1 results carried out with all five HERV primer Pairs on Human 293T DNA. Results of PCRs carried out on human 293T DNA. The wells labelled with "**H1**" to "**H5**" are PCR results with the primer pair number indicated. The wells labelled with "**-**" are negative control wells containing the PCR results which were obtained without adding target DNA to show if the primers interact with each other in any way. The white ovals indicate the near 7,500 bp fragments expected. The 1KB ladder was used to compare the sizes of bands and is indicated on the right. The gel image was captured using the molecular imager gel doc XRS.

As seen in **figure 5.2**, all five pairs, **H1 – H5**, contain a band at the expected size of around 7,500 bp as indicated by the white ovals. Bands are seen in the negative control well, **H1-**, as well as the adjacent empty well to the right of the **H1-** indicating that there could have been spill over of genetic material from the **H1** well during gel loading. **H5-** also shows bands, indicating primer dimerization or spill over. None of the other negative control wells, **H2-**, **H3-** or **H4-** indicated any bands as would be expected since no DNA was added. PCR results from **H1** and **H5** were discounted due to unreliability. All primer pair products indicated bands at non-expected sized,

indicating non-specificity of the primers. Further repeat PCR reactions were carried out to obtain clearer results. LTR primers HL1 – HL3 were also run.

5.2.2 Repeat PCR along with LTR primers
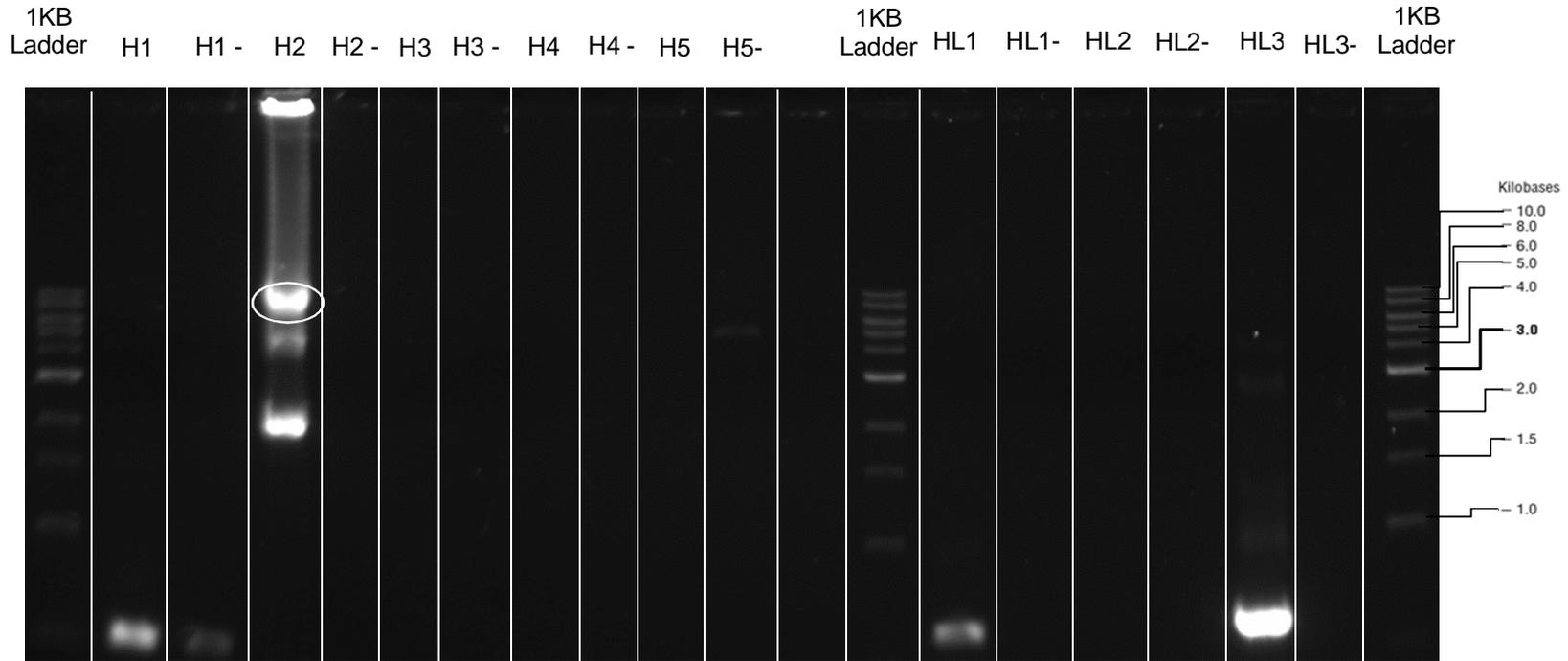
Figure 5.3 Gel image of PCR-2 results carried out with five HERV primers and three LTR primers on 293T DNA. Results of PCRs carried out on human 293T DNA. The wells labelled "**H1**" to "**H5**" are PCR results with the primer pair number indicated. The wells labelled "**HL1**" to "**HL3**" are primer pairs which attach to the LTR regions of the HERV provirus. The wells labelled with "-" are control wells containing PCR results obtained without adding target DNA to show any inter-primer interaction. The white oval indicates the 7,500 bp fragment expected. The 1KB ladder was used to compare the sizes of bands and is indicated on the right. The gel image was captured using the molecular imager gel doc XRS. White lines have been overlaid over the image to distinguish separate wells.

As seen in **figure 5.3** following a repeat of the PCR carried out in **figure 5.2** along with the LTR primers, only **H2** shows a band at the expected size of around 7,500 bp as indicated by the white oval. **H5** shows a band at the wrong size at around 5,000bp, a band commonly seen across all primer reactions as seen in **figure 5.2**. This is further indication of primer non-specificity. **HL3** is the only LTR primer to show any product, however, these bands not the expected size of around 8.500 bp. None of the negative controls have any bands indicating primers are not interacting with each other; this further indicates that the bands seen in the negative control wells in **figure 5.2** were more likely to have been caused by produce spill over from adjacent wells during loading.

As the LTR primers did not produce any results around the expected size, it was decided to focus on optimising the original **H1**-**H5** primers. PCR was also attempted on non-human DNA to establish that they only function to amplify human DNA.

5.2.3 HERV PCR repeats on 293T and V79 DNA

A - Low Exposure Result
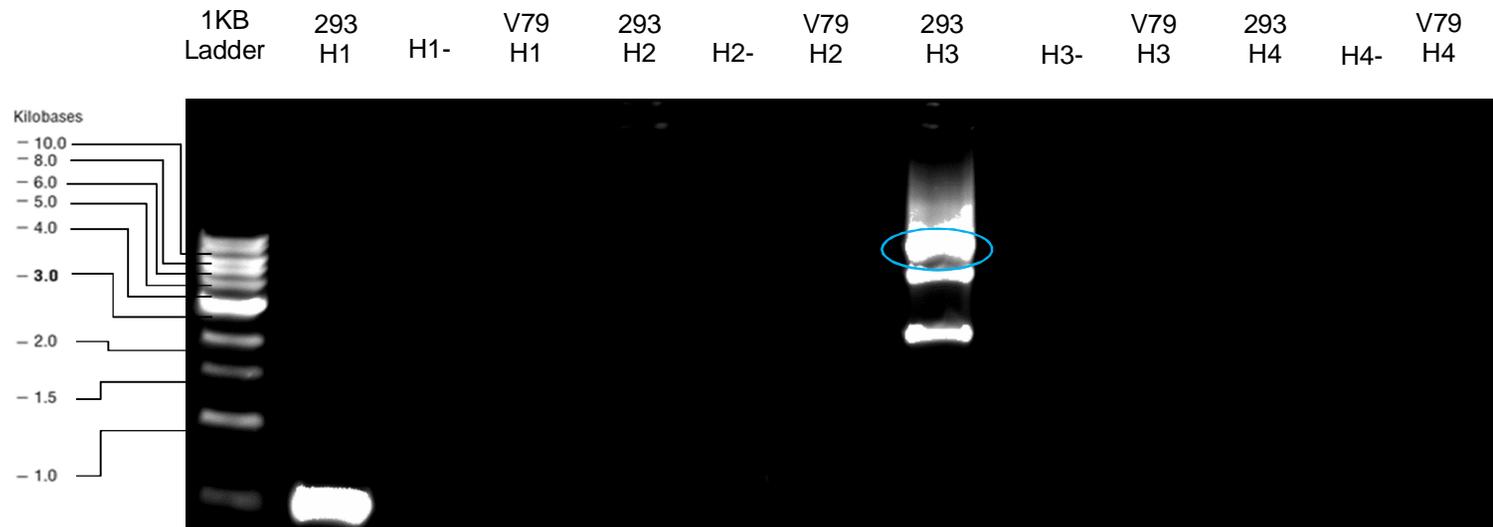
B - High Exposure Result
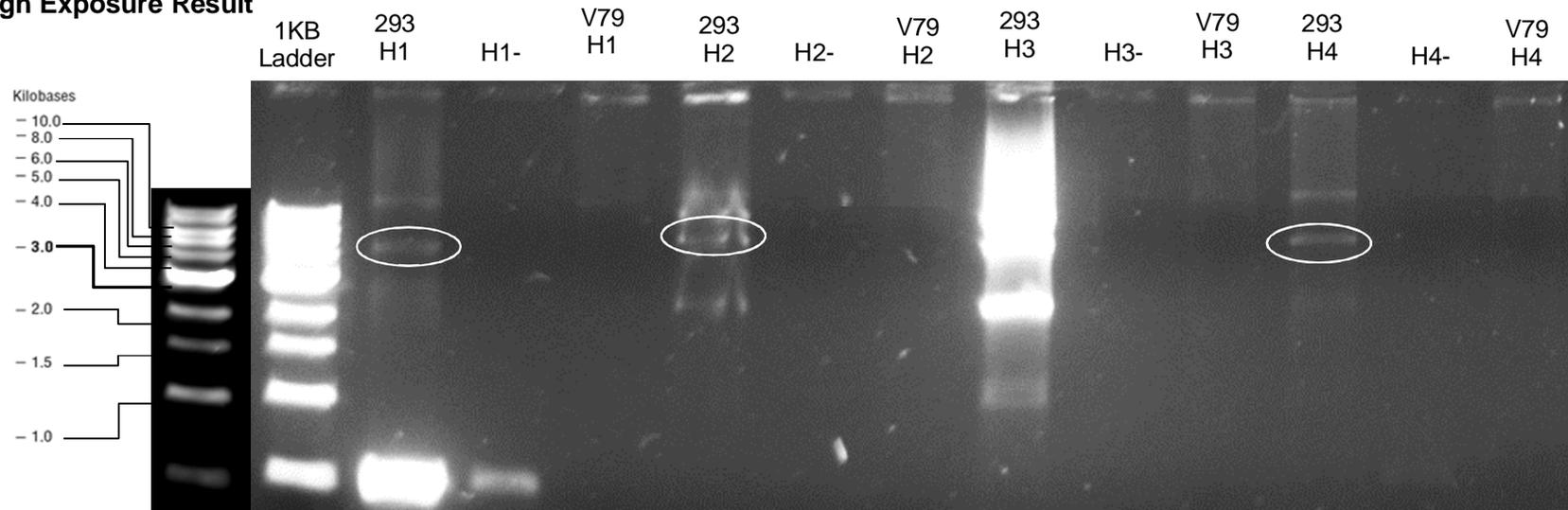
Figure 5.4 Gel image of PCR-3 carried out with 4 HERV primers on 293T and V79 DNA at low and high exposure. Results of PCRs carried out on human 293T DNA and Chinese hamster V79 DNA. **A**) shows the results at low exposure, while **B**) shows the results at high exposure. Wells are labelled with the cell line DNA being acted on and the primer pair being used. The wells labelled "H1" to "H5" are PCR results with the primer pair number indicated. Wells labelled with "293" represent the 293T human DNA while those with "V79" contain V79 Chinese Hamster DNA. **A**) was imaged at low exposure to try and discern the bands in the 293 **H3** well which amplified a very high amount of PCR product. **B**) was imaged at low exposure to observe the bands in the results outside the 293 **H3** well. The ladder from the low exposure image (**A**) is shown alongside to more clearly determine the band sizes. The white ovals indicate the near 7,500 bp fragments expected. 1KB ladder was used to compare the sizes of bands and is indicated on the left. The gel image was captured using the molecular imager gel doc XRS.

In **figure 5.4 A**, the blue oval within the **293 H3** well indicates the area which may contain the near 7,500bp band which is expected from the **H3** primers. Other bands are seen, still indicating a lack of primer specificity. A high intensity band is also seen in the 293 H1 lane at the 1,000bp mark, which is not expected. A high intensity band was seen in the H5- control well, indicating an error, thus the H5 could not be counted and so the H5 PCR results are not shown.

115

In **figure 5.4 B**, The white ovals indicate the area which may contain the near 7,500bp band which is expected from the primers. No bands are seen in the control wells labelled **H1-** to **H4-** which shows that no inter primer reactions are occurring as there is no template DNA for the primers to act on. No product is seen in the Chinese hamster V79 wells either, showing that the primers are indeed specific to human DNA which is where the HERVs will be present. Other bands are seen along with the 7,500bp band; indicating a potential lack of primer specificity much like prior results seen in **figures 5.2** and **5.3**.

The PCR was then repeated using the **H3** primers on the human 293T DNA, Chinese hamster V79 DNA and mouse 3T3 DNA. The **H3** primers were selected due to the very high amount of PCR product they consistently produced as seen in **figures 5.2** and **5.4**. The non-human V79 and 3T3 DNA was selected confirm that the primers only acted on human 293T DNA, where the HERV sequences are present.

5.2.4 HERV PCR with 293 DNA, V79 and 3T3 DNA
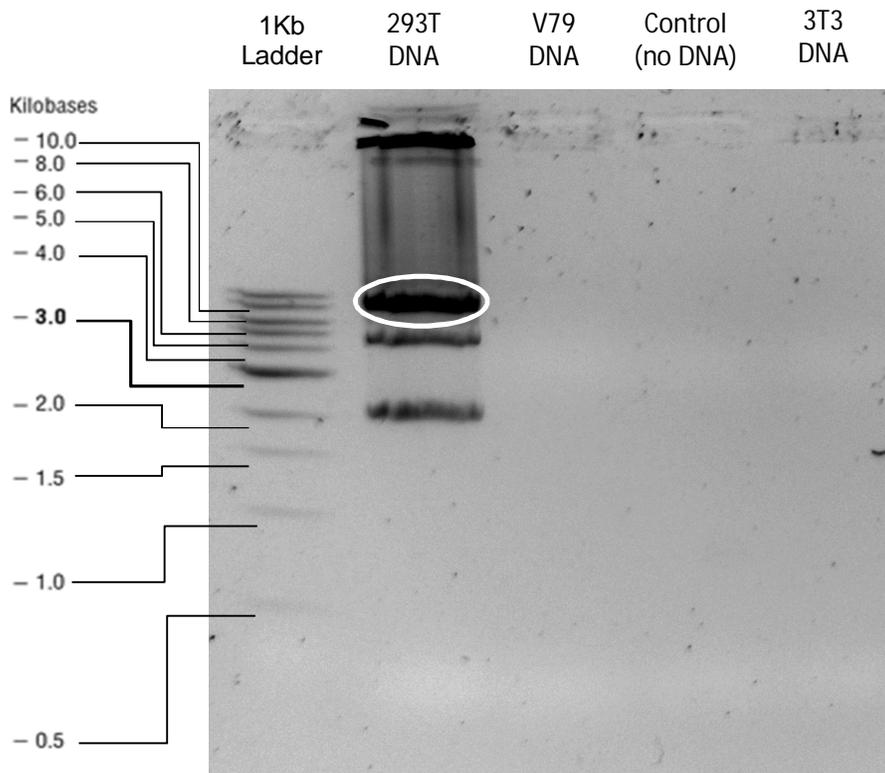
Figure 5.5 Gel image of PCR-5 carried out with HERV primer pair 3 on 293T, V79 and 3T3 DNA.
PCR results **H3** primers on 293T human cells, V79 Chinese hamster cells and 3T3 mouse cells. The white oval indicates the expected 7,500bp fragment. The 1KB ladder was used to compare the sizes of

bands and is indicated on the left. The gel image was captured using the molecular imager gel doc XRS.

As seen in **figure 5.5**, PCR product only seen in the 293T well. The white oval indicates the approximately 7,500bp desired band. No bands are seen under the V79 and 3T3 wells confirming that the primers only function on human target DNA. This further corroborates that the PCR is specific for a HERV sequence that is only present in human cells. However, the superfluous bands are seen again in the 3T3 well.

A step towards optimising the primer specificity was taken by attempting the PCR with Q solution provided by the manufacturers along with the Long Range PCR kit from Qiagen. Q solution contains proprietary chemicals purported to be able to improve or inhibit primer specificity and the manufacturers have suggest to attempt using the solution to possibly improve results. This information is available in the Long Range PCR kit from Qiagen.

5.2.5 HERV Repeat with Q solution
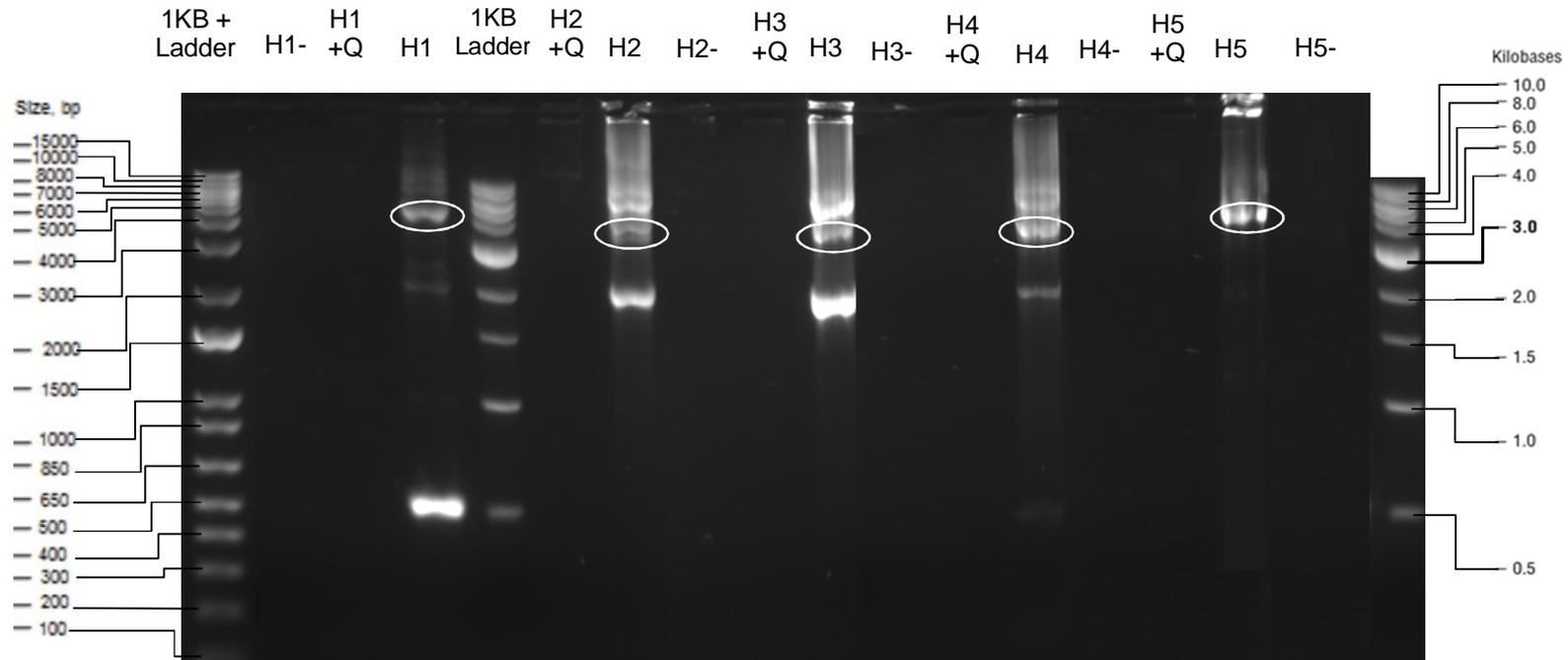
Figure 5.6 Gel image of PCR-6 carried out with five HERV primers on 293T DNA with Q-solution. Results of PCRs carried out on human 293T DNA. The wells labelled "**H1**" to "**H5**" are PCR results with the primer pair number indicated. The wells marked with "**+Q**" indicate the use of Q-solution in the primer mixture. The wells marked with "**-**" indicate negative controls without any target DNA to be acted on. The white ovals indicate the expected 7,500 bp band. The 1KB and 1KB+ ladders were used to compare the sizes of bands. The 1KB+ ladder are indicated on the left of the gel electrophoresis image while the 1KB ladder is shown on the right side of the image. The gel image was captured using the molecular imager gel doc XRS.

As seen in **figure 5.6**, **t**he Q-solution was seen to completely inhibit the PCR from occurring, while all five primer pairs were able to produce the expected 7,500bp band without Q-solution as indicated by the white ovals. All five pairs acting without Q-solution, once again, showed superfluous bands.

Q-solution did not improve the results as it seemingly prevented any PCR reaction from taking place. As such, it was decided to move on to generating a large scale PCR to generate a large amount of PCR product for further analysis. Primer pairs **H2**, **H3** and **H4** were chosen for this as they produced the fewest non-specific bands; with **H3** also producing the strongest bands.

5.2.6 Large Scale HERV PCR Attempt 1
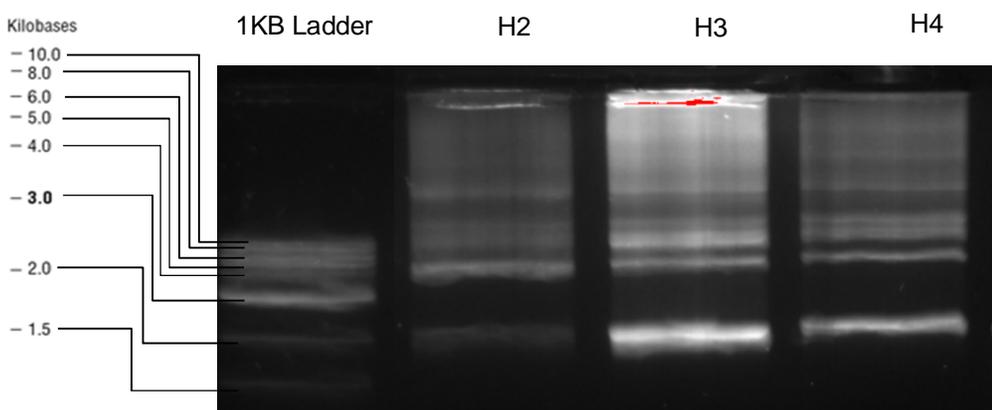
Figure 5.7 Gel image of large scale PCR-7 carried out with three HERV primers 293T DNA. PCR results of a large scale PCR carried out on human 293T DNA. A Large amount of PCR product was attempted to be produced in larger wells with a higher volume of 100 μl as opposed to the standard 50 μl reaction as indicated in **section 2.2.17**. The wells labelled **H2** - **H4** indicate the primers used for the PCR. The 1kb ladder is shown on the left. The gel image was captured using the molecular imager gel doc XRS.

As seen in **figure 5.7**, all three primer pairs, **H2**-**H4**, produced results. However, the bands all had a very streaked appearance. The streaked nature of the bands made them unsuitable for the next part of the experiments. As such the large scale PCR was repeated with less initial 293T DNA to be amplified.

A second attempt was conducted to produce more defined bands: PCR 8. This PCR showed no amplification at all, indicating that the 293T DNA sample that was being used was too little an amount for the PCR to work. More 293T DNA was then extracted before further repeats were performed to determine the correct amount of the 293T DNA required to produce clear bands that were not streaked.

119

5.2.7 Large Scale HERV PCR Attempt 3
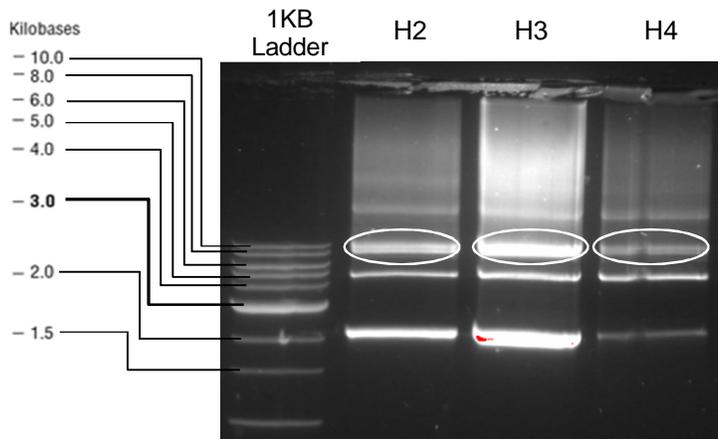
Figure 5.8 Gel image of PCR-9 carried out with three HERV primers on 293T DNA. PCR results of a large scale PCR carried out on 293T DNA. A Large amount of PCR product was attempted to be produced in larger wells with a higher volume of 100 µl. The wells labelled **H2** to **H4** indicate the primers used for the PCR. The white ovals indicate the position of the 7,500bp bands. The 1kb ladder is shown on the left. The gel image was captured using the molecular imager gel doc XRS.

As seen in **figure 5.8**, clearer bands were seen compared to **figure 5.7**. **H3** showed the strongest signal. The expected 7,500 bp bands are indicated with the white ovals; these were then cut out from the gel and purified as described in **section 2.2.17**. The bands were run in a new gel to determine if they were of the correct size.

5.2.8 HERV band removal and purification
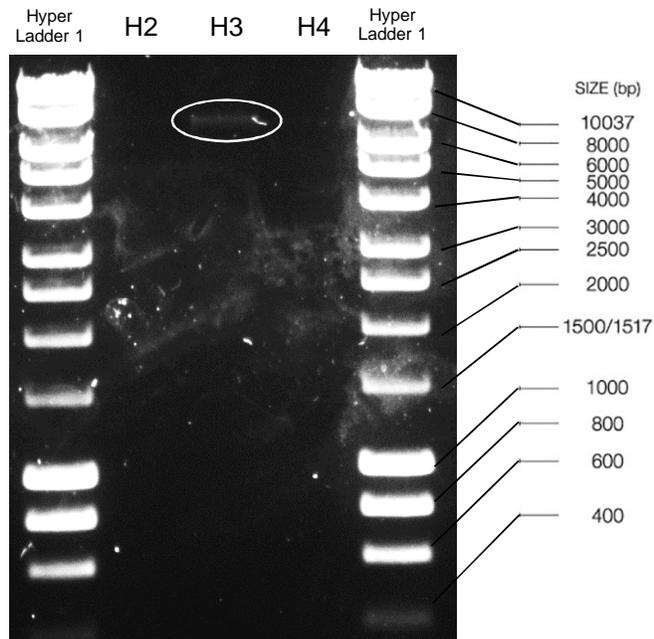
Figure 5.9 Gel image showing results following band purification from PCR 9. Following extraction and purification of the target bands in **figure 5.8**, a sample was run to show successful extraction. The wells labelled **H2** to **H4** indicate the primers used for the original PCR. The white oval indicates the **H3** band. DNA Hyper Ladder 1 is shown on the right with band sizes. The gel image was captured using the molecular imager gel doc XRS.

As seen in **figure 5.9**, a clear, yet faint band is seen around the 7,500 bp mark for **H3,** however, the **H2** and **H4** bands did not appear to further optimise the HERV PCR to produce a stronger band from the **H2** and **H1** primers, their forward and reverse primers were paired to try and find a more optimal primer pairing. This "primer mixing" was also paired with lowering the annealing temperature to attempt to increase primer specificity.

5.2.9 HERV PCR Primer Pair Optimisation
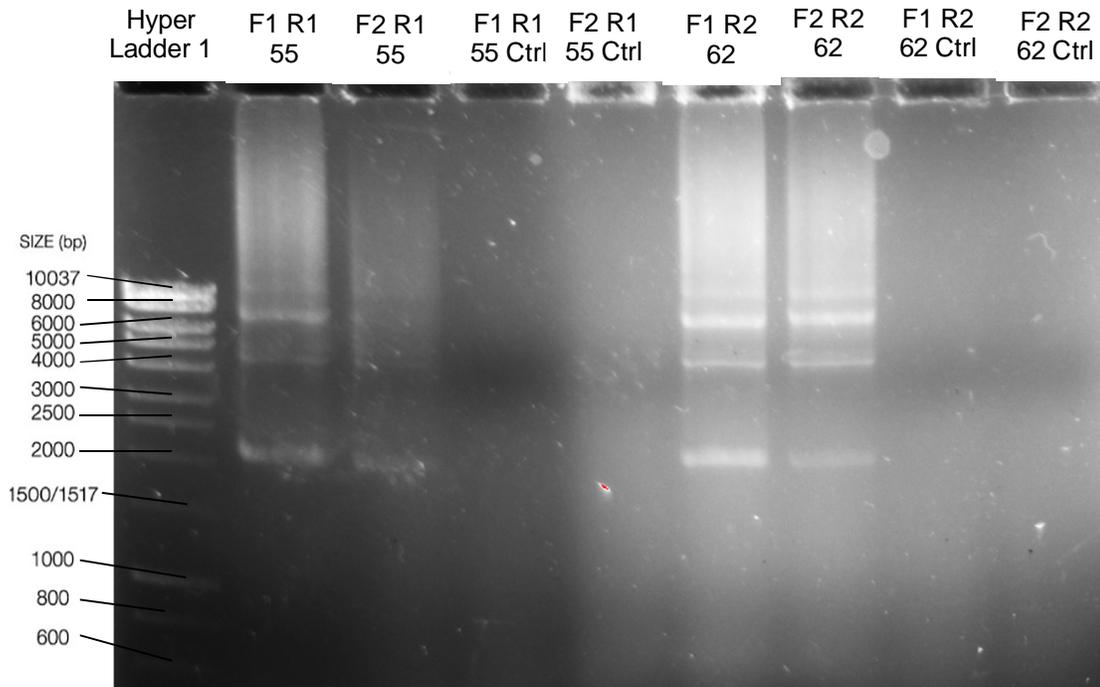
Figure 5.10 Gel image of PCR-10 with primer and temperature adjustment using H1 and H2 primers. 293T DNA was amplified. The wells indicate which forward and reverse primers were used. **F1** and **F2** indicate the forward primer from HERV primer pair 1 and 2 respectively. **R1** and **R2** indicate the reverse primer from HERV primer pair 1 and 2 respectively. The wells marked 55 and 62 indicate the annealing temperature the PCR reaction was performed at 55°C and 62°C respectively. The R1 primers were run with lower annealing temperatures due to their lower GC content compared to the R2 primers. The wells marked "**ctrl**" are the control wells lacking the target DNA to test for primer dimerization. DNA Hyper Ladder 1 is shown on the left with band sizes. The gel image was captured using the molecular imager gel doc XRS.

As seen in **figure 5.10**, the use of the lower annealing temperature of 55°C paired with the R1 reverse primers (due to a lower GC content as compared to the R2 primers) instead of the annealing temperature of 62°C along with the R2 primers produced lighter bands. This shows that the R2 primers work better. The strongest bands were seen with the F1 and R2 primers and the F2 and R2 primers. Since the intensity of the bands did not appear to improve significantly, it was decided to work with the H3 primers as they produced the strongest bands consistently. Further large scale PCR amplification was done with the H3 primers to generate a large amount of PCR product for restriction digest profiles.

5.2.10 HERV Gel Image 10 - HERV Band Purification repeat
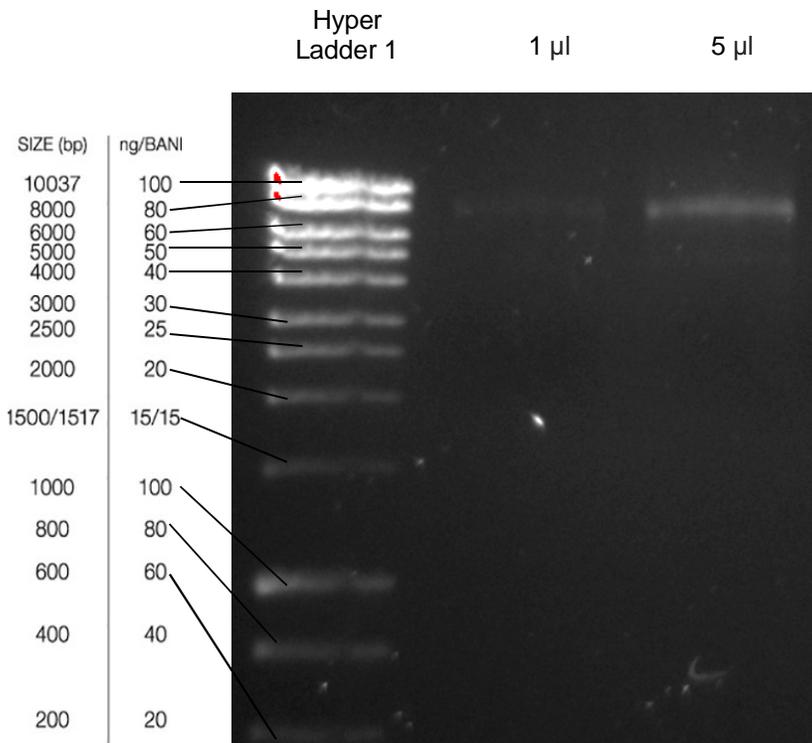
Figure 5.11 Gel image showing results following a second band purification from PCR 9. After more PCR product from **figure 5.8** was run in a gel, the 7,500 band was cut out and run in the gel seen. 1 μl and 5 μl of the sample were loaded as shown. Hyper Ladder 1 was used to determine band size. The gel image was captured using the molecular imager gel doc XRS.

In **figure 5.11**, the wells are marked with the amount of PCR product loaded into each well. A clear band is seen below the 8,000bp mark, around where the expected 7,500bp band should be. 5 μl and 1 μl samples were loaded into two wells in case the 1μl sample was faint as was the case in **figure 5.9**. In the case of **figure 5.11**, the 1 μl sample was indeed faint, but the 5 μl confirms the existence of the band. This is still faint compared

to the initial brightness seen by the H3 primer pair PCR seen in **figure 5.8**, thus it can be inferred that, following the band being cut out from the gel, the purification process causes some loss of the band.

Once enough of a PCR product was obtained, a restriction digest was planned using the website NEBcutter (Vincze, Posfai and Roberts, 2003), a program which analyses DNA sequences for sites that can be cut by restriction enzymes and can also show an expected gel electrophoresis result following digestion.

5.3   Restriction Digest

5.3.1   Digest Planning

The part of the HERV K113 sequence (Turner *et al.*, 2001) that is amplified by the H3 primers was input into NEBcutter (Vincze, Posfai and Roberts, 2001) to determine which restriction enzymes can be used to identify the sequence. Of the restriction enzymes that could be used EcoR32I (EcoRV) and KpnI were determined to be useful as they only made a single cut each and were available for use.
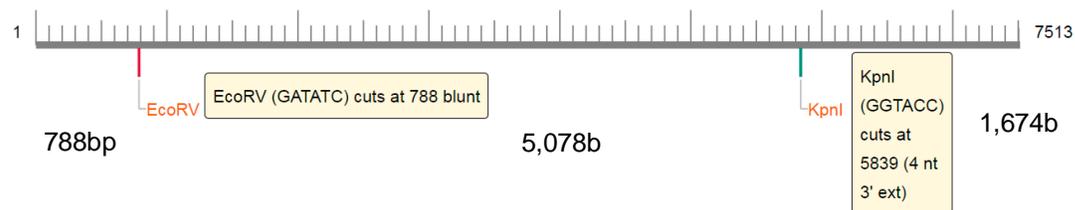
Figure 5.12 Restriction Digest profile of the H3 product showing where the EcoR32I and KpnI cut. EcoR32I will cut the product at the 788bp mark and KpnI will cut it at the 5,839bp mark. This will cut the original 7,513bp sequence into three parts of sizes 788bp, 1,674bp and 5,051bp. The image was generated by the NEBcutter website.

NEBcutter also provided an example of what the restriction profile can look like as seen in **figure 5.13** below.
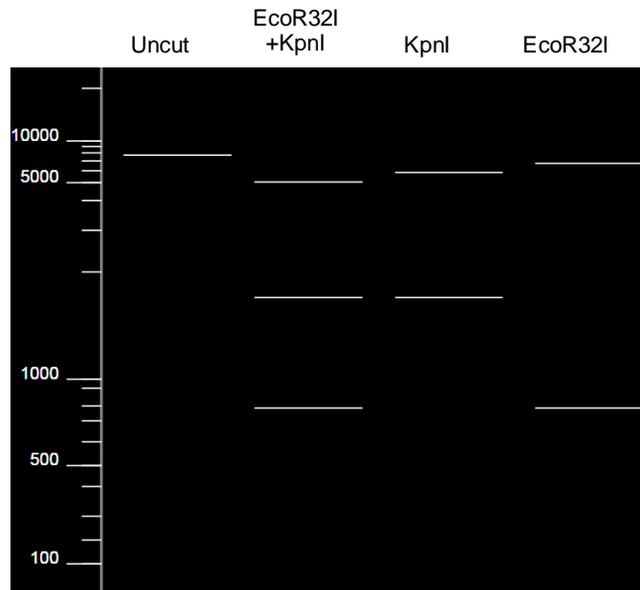
Figure 5.13 Expected gel profile following H3 PCR product digestion with EcoR32I and KpnI. The labels above the wells indicate the lanes with the uncut 7,513 bp band, H3 PCR product following restriction by individually by EcoR32I and KpnI, as well as both of them. The uncut band is seen on the left at 7,513 bp. KpnI produces two bands at 5,839bp and 1,674bp. EcoR32I produces two bands at 6,725bp and 788bp. Both enzymes will produce three bands at 5,051bp, 1,674bp and 788bp. Image was generated using the NEBcutter website.

5.3.2 Restriction Digest

In order to ensure the functionality of the restriction enzymes, can also be used to linearise plasmids that they are known to have a cutting site for. KpnI is known to cut plasmid p54 (Rosonina *et al.*, 2005) and EcoR32I is known to cut plasmid p68 (Scherbakova and Verkhusha, 2013).

An intact circular plasmid will run further in agarose gel due to its compact nature; once cut by a restriction enzyme and linearised, the DNA will not be able to move as freely through the agarose gel. Thus, if the plasmid moves a shorter distance when a restriction enzyme is used, the restriction enzyme is known to be functional.

Restriction digestion was performed as described in **materials and methods section 2.2.18**.
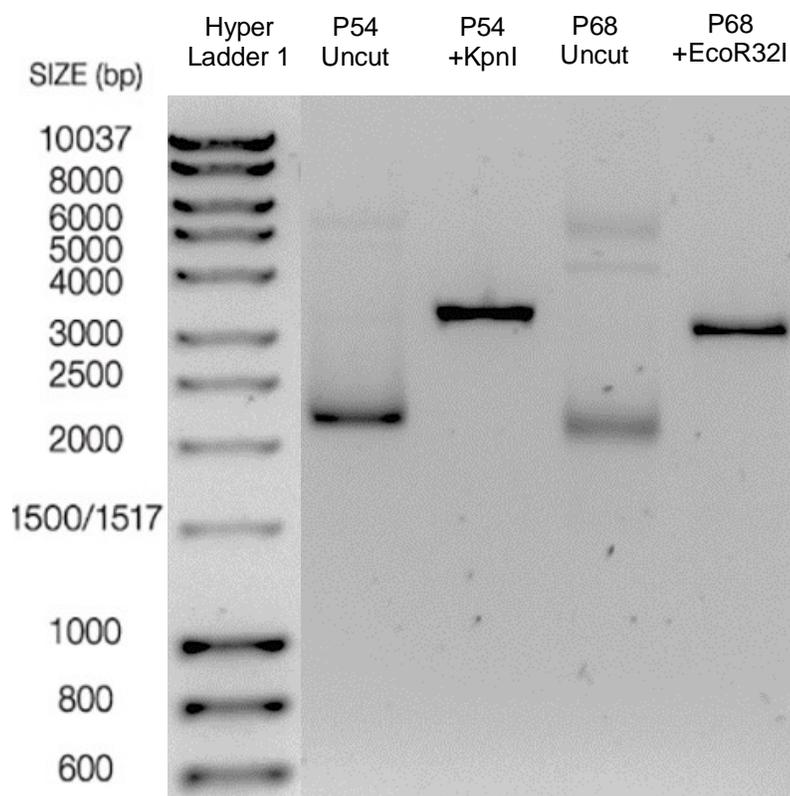
Figure 5.14 Gel image of restriction digestion of plasmids p54 and p68. The circular plasmids, p54 and p68, indicated as uncut, have run a longer distance in the agarose gel compared to the linearised versions which have been digested with KpnI and EcoR32 respectively. The Hyperladder band sizes are indicated next to the ladder well. DNA Hyperladder 1 is indicated with band sizes on the left. The gel image was captured using the molecular imager gel doc XRS.

As seen in **figure 5.14**, due to their compact nature, the circular plasmids are able to pass more easily through the agarose gel. Once they have been linearised, the bands travel a shorter distance following incubation with the restriction enzymes. This indicates that the restriction enzymes were functional.

Once restriction enzyme functionality was confirmed, the H3 PCR band was digested and the result was run in a gel as seen below.
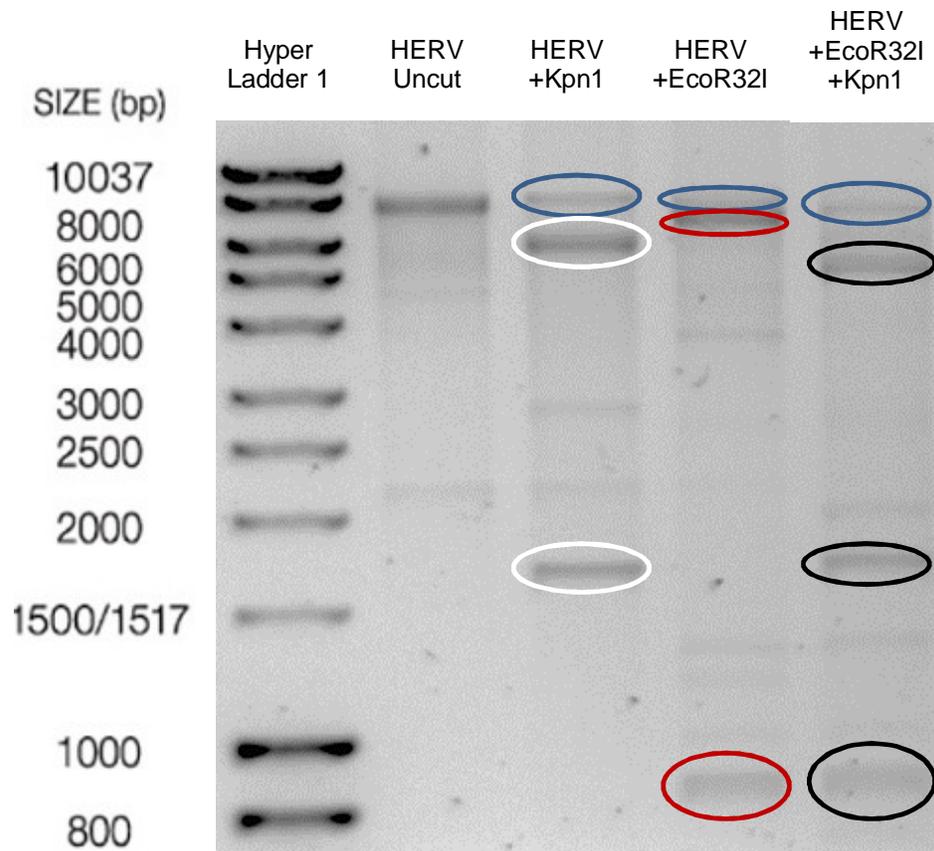
Figure 5.15 Gel image following restriction digests of the H3 PCR product. Results following restriction digest of the H3 PCR product. The uncut HERV well contained the 7,513 bp product. The Other three wells have the restriction digest indicated. The blue ovals indicate remnants of the uncut band in the restriction digest wells. The white ovals show the expected KpnI products. The red ovals show the expected EcoR32I products and the black ovals show the expected the products with combined use of KpnI and EcoR32I. Other faint unexpected bands are also seen. DNA Hyperladder 1 is indicated with band sizes on the left. The gel image was captured using the molecular imager gel doc XRS.

As seen in **figure 5.15**, the uncut HERV band is at the expected mark of around 7,500bp.The blue ovals indicate that within all three wells which utilised the restriction enzymes, the uncut HERV is still present along with other bands, indicating that not all the PCR product was completely digested thus the restriction could be further optimised. The ovals indicate the bands expected after restriction as shown in figure seen in **figure 5.13**. The white ovals show the expected KpnI bands at 5,839 bp and 1,674 bp. The red ovals show the expected EcoR32I bands at 6,725 bp and 788 bp. The black ovals show the bands both restriction enzymes were expected to produce at 5,051 bp, 1,674bp and 788bp. Other faint unexpected bands are also seen. The presence of unexpected bands shows that the H3 PCR sample cut out from the gel in **figure 5.8** may have contained

the other superfluous bands seen. This theory is also corroborated by the uncut HERV also showing bands that should not be present.

While there were non-specific bands seen following restriction digestion, the presence of the expected bands (shown in the ovals) indicates that the original uncut HERV was present in the H3 PCR product.

5.4 Discussion

5.4.1 General Results

Primers were designed to amplify a near 7,500 bp sequence of HERV-K113 to determine if the presence of said HERV could be detected in genomic DNA extracted from cells.

The PCR primers, while non-specific, were able to produce a band of expected length of around 7,500bp when used to amplify human 293T DNA as seen in **figures 5.2** to **5.11**. The primers produce no bands when used on non-human V79 and 3T3 DNA as seen in **figures 5.4 – 5.5**. That the sequence amplified was only found in human cells indicates that the targeted sequence is indeed only present in humans, lending credence to the theory that the amplified sequence was indeed from a HERV sequence as HERVs are thought to be evolutionarily tied to humans (Feschotte and Gilbert, 2012; Dewannieux and Heidmann, 2013) and would not be present in non-human cells.

The restriction digestion experiments also served to prove that the PCR product was indeed the one expected, since the exact expected band profile seen when comparing the actual results (**figure 5.15**) and the expected results (**figure 5.13**). Although, while the expected bands are seen, there are still unexpected bands seen in the restriction gel image (**figure 5.15**). These include one seen in the uncut HERV well, indicating that the purification of the HERV product following band extraction needs to be improved. These superfluous bands that are seen in the uncut HERV well may be also having sites that the restriction enzymes utilised can cut.

The superfluous bands formed may be due to parts of the HERV-K113 virus being present in incomplete parts across the human genome (Wildschutte *et* al., 2016), thus causing amplification of segments not of the expected size. This would mean that chunks of the HERV sequence with the correct binding sites could be interacted with by the primers, leading to the amplification of shorter sequences not containing the entire HERV sequence. This concept is demonstrated below, in **figure 5.16**.
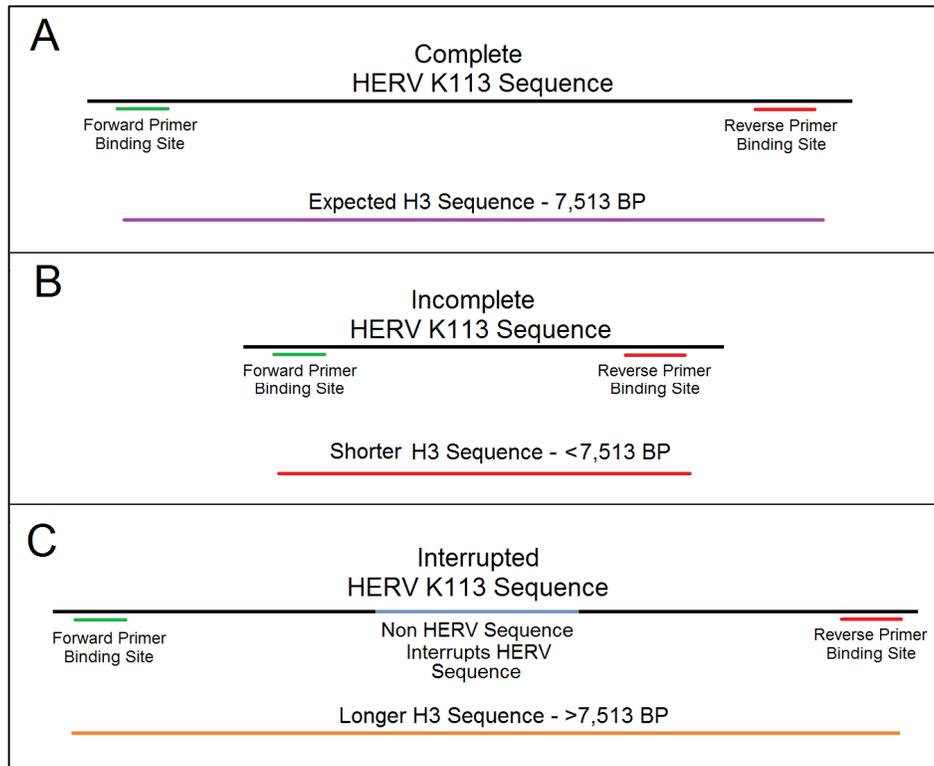
Figure 5.16 Diagram Showing the Incomplete and Interrupted HERV K113 Sequences. **A**) Shows the primers binding to the complete HERV K113 sequence and producing the expected result. **B**) Shows the primers binding to an incomplete HERV sequence, leading to the formation of a shorter PCR product than expected. **C**) Shows the HERV sequence interrupted by other sequences, leading to the primers forming a longer than expected PCR product. The diagram was made using Microsoft Paint.

Figure 5.16 shows how the spread of the HERV K113 sequence in patches across the human genome can lead to primer binding sites attaching to non-intact HERV sequences, leading to the formation of superfluous PCR product. Figure **B** could explain the formation of the shorter bands seen across the HERV PCRs. Limitations of the LongRange PCR kit might be preventing the amplification of bands significantly larger than 7,500 bp that would be formed in the situation figure **C** indicates.

The likelihood that the HERV band produced is only comprised of one or many coincidental sized band at around 7,500 bp is quite low due to the fact that the exact pattern of expected HERV bands is seen for both restriction enzymes (figures **5.13** and **5.15**). There is, however, a chance that, in addition to the expected HERV sequence, there are other sequences also being amplified around the 7,500 bp mark. This idea is corroborated by the restriction digest results seen in **figure 5.15**; in this figure the wells which underwent restriction show the presence of the uncut near 7,500 bp band (indicated with the blue ovals). The presence of a band around the expected uncut

HERV sequence may mean that the restriction digest could be further optimised, but there is also a chance that there are bands which are coincidentally the size of the HERV band at around 7,500 bp.

In order to better understand the products produced by the HERV 3 primer pair, a primer BLAST (Ye *et al.*, 2012), was conducted with the H3 primer pair on the human genome; these results are shown in detail under **Appendix 5**. Primer BLAST can be used to provide potential products to be expected on a specific target sequence. While several PCR products were predicted, there were 20 that fit the near 7500bp size range, with HERV sequences accounting for 7 of them. This shows that the primer specificity could have been improved.

5.4.2 Further Work

Primer design for primers could be refined, with more extensive use of the primer BLAST site to design more specific primers.

The process of purifying the suspected HERV band after being cut out from the agarose gel can be optimised to prevent any superfluous bands from also being removed. The restriction digest profile can be further optimised if necessary following an optimised PCR.

The HERV PCR can be further optimised to only produce the desired band, though this may not be possible if incomplete portions of the provirus are being amplified. The restriction digest profile can be further optimised if necessary following an optimised PCR and an optimised target band removal from agarose gel following electrophoresis. Though this may not be possible, once again, if incomplete portions of the HERV-K113 provirus are being amplified which are close in size to the expected band.

Due to time constraints, experiments to determine if human produced empty vectors could transfer HERVs to non-human cells could be performed. To perform these experiments, non-human cells, such as V79 or 3T3 cells, would be infected with empty vectors produced by human cell lines. If the HERV sequence is found in the non-human cell lines following infection with human produced empty vectors, then there is an indication that the HERVs are being packaged into the empty vectors and integrated into the host cells.

HERVs such as HERV-K113 are known to be capable of producing intact viral particles (Boller *et al.* 2008) and it is known that empty vectors can form as part of a mixture of

gene therapy vectors (Zhao et al., 2008). If these vectors are able to package genetic material such as HERV K113, they might cause insertional genotoxic damage. Batches of retroviral vectors could be sequenced for the genetic material carried by them; this would reveal if any have packaged the HERV sequences. The detrimental capacity of these HERV carrying vectors could be tested, primary in the form of integration related genotoxicity.

A method to confirm that the band contained the correct HERV sequence involves sequencing the band for all the genetic sequences present (Dorit *et al.*, 2001). This would allow a confirmation the HERV sequence from the superfluous sequences.

HERVs, such as HERV-F have been found using hybridisation and Southern blotting (Widegren *et al.*, 1996; Kjellman *et al.* 1999). A similar method could be used with the planned PCR protocol described in this chapter. The Southern blot test would be established and used to isolate the HERV K113 genome from human genomic DNA. Following this, human produced retroviral vectors could be used to infect non-human cells, then those non-human cells could be tested with the Southern blot protocol to establish if HERV are being taken up by retroviral vectors and being integrated into host cells.

HERV Sequences, such as HERF-H sequences, have previously been analysed for specific features, such as HERV-H specific ORFs, *gag*, *pol* and *env* sequences and LTR sequences, which were then utilised *in silico* and compared against existing human genomic databases to locate entire full length HERV-H sequences in the human genome (Jern *et al.*, 2005). Other researchers have used *in silico* methodologies, such as the use of RNAseq datasets with more sequence depth (more reads per sample), to search for elements close to entire HERV provirus sequences to locate locus specific HERVs (Hamann, Abida and Lange, 2023). Methodologies such as these could be used to locate more full length HERVs, which could then be analysed for genotoxicity.

Chapter 6 - Discussion

6.1 Background

Gene therapy is a medical technology which aims to treat or prevent a disease caused by a defective or non-functional variant of a gene. The ideal use of gene therapy to treat a condition allows it to repair or replace a defective genetic sequence, cause the therapeutic effects to function over a long period, while also conserving the location of the correct gene in its natural location while not causing any disruptions to the regulatory material around the altered gene, thus preventing any genetically induced side effects (Crick, 1970; Morange, 2009; Mammen, Ramakrishnan and Sudhakar, 2007; Medline, 2022).

One method of gene therapy involves the use of viral vectors, which are naturally optimised to incorporate genetic sequences into cells they infect as a part of their life cycle (Gillet *et al.*, 2009). The retroviruses (RVs) and lentiviruses (LVs) are examples of such viruses which can transfer genetic information directly into the genome of host cells and allowing permanent expression which is retained following cell division (Durand and Cimarelli, 2011; Skalka and Katz, 2005; Miller, Farnet and Bushman, 1997).

The use of LVs or RVs as vectors for gene therapy involved several generations of alterations to their biology in order to render them unable to be harmful to the host cells (Zufferey et al., 1998; Escarpe et al., 2003). Ultimately, the goal of using RVs and LVs in gene therapy is to use them as vectors for delivering therapeutic genes which can be into host cells without causing harm functions by replacing the viral genome with a therapeutic gene (Zhao et al., 2008; Gillet et al., 2009).

However, RV gene therapy has been liked to severe side effects, including oncogenesis as seen with leukaemia formation in animal models and in human patients during the X-SCID gene therapy trials in the UK and France (Cavazzana-Calvo and Hacein-Bey-Abina, S., 2001; Cavazzana-Calvo *et al.*, 2000; Gaspar and Thrasher, 2009). In research carried out following the X-SCID trials, it was determined that promoters and enhancers in the RV vectors upregulated *LMO-2*, a gene known to be involved in the causation of leukaemia in children (Bjorgvinsdottir *et al.*, 1997; Stein *et al.*, 2010; Malech *et al.*; 1997 Gaspar and Thrasher, 2009). Gene therapy has also caused myelodysplasia in patients undergoing gene therapy for chronic granulomatous disease (Stein *et al.*, 2010; Malech *et al.*, 1997; Bjorgvinsdottir *et al.*, 1997). RV gene therapy has also

caused genotoxicity via insertional mutagenesis. This genotoxicity has been seen to cause the down-regulation of tumour suppressor genes and chromatin remodelling, both of which can lead to tumour development (Lazo and Tsichlis, 1988; Ben-David *et al.,* 1990). Considering all the possible side effects that RV gene therapy can cause, it is important to understand some of the ways in which the RV vectors can cause damage to the host.

As part of their life cycle, RVs form double strand breaks (DSBs) in the host genome during integration to form a site insertion their reverse-transcribed viral genome DNA into the host cell (Skalka and Katz, 2005). The RV then relies on the host cell non-homologous end joining (NHEJ) mechanism to then repair the DSB, incorporating the viral genome within the host DNA (Skalka and Katz, 2005). NHEJ, unlike homologous recombination, is an error prone process which can cause the loss of genetic material as the process does not attempt to conserve the genetic sequence across broken strands by the use of a reference strand (Mohiuddin and Kang, 2019). Indeed, the lack of use of a reference is what allows the viral DNA to be incorporated into the host cell, since any new viral sequence would be detected by the use of reference strand and simply be removed (Skalka and Katz, 2005; Miller *et al.*, 1997; Mao *et al.*, 2008). It is also this lack of reference which allows us to replace the viral genome in a RV vector with a therapeutic sequence, which will then not be removed from the host cell. However, this inherent erroneous nature of NHEJ can cause conditions such as Burkitt's lymphoma (Rowh et al., 2011).

The DSBs themselves can be oncogenic, and have been shown to cause chromosomal aberrations, rearrange the host genome and can even be cytotoxic to the point of causing ell death via apoptosis (Smart *et al.*, 2008). Thus, cells with impaired NHEJ mechanisms can be at worse risk for severe damage to their DNA. If a retroviral vector is used on NHEJ impaired host cells, the DSBs formed during integration will not be adequately repaired. As such, it is necessary to determine the functionality of the NHEJ mechanism of a gene therapy candidate before being targeted for RV gene therapy.

Work carried out for this thesis could be developed into an assay to determine if RV gene therapy is a viable option for gene therapy candidates.

Work carried out in **chapter 3** involved the analysis of cells for DSB formation with a γH2AX assay utilising immunocytochemistry (ICC). H2AX is a histone protein which becomes phosphorylated into γH2AX. The assay presented in **chapter 3** is able to produce a DSB profile over 48 hours. Individuals whose cells had increased DSB

formation following RV vector transduction by the 48 hour mark with a poor DSB repair profile could then be singled out as candidates with a potentially impaired NHEJ mechanism.

An assay could be formed which would allow gene therapy candidates to provide a cell sample which could be infected with the retroviral vector and analysed for their DSB profile. In 48 hours (along with a few hours' time for analysis) these individuals could then be advised appropriately before using RV vector gene therapy.

The assay could be improved by utilising other methods of assessing DNA damage with a faster analysis time. For example, the ImageStream analysis throughput can be vastly faster than ICC, reducing the time taken for analysis by hours. Other methods such as comet assays or specialised ELISA models could be developed for these assays to provide more cost effective methods of analysis compared to ImageStream.

While research has been carried out on the effects of DSB formation on cells (Albanese *et al.*, 2008; Smart *et al.*, 2008; Baum *et al.*, 2004; Mitchell *et al.*, 2004; Wu *et al.*, 2003; Schröder *et al.*, 2002), no such assay has been developed to determine the functionality of the NHEJ mechanism. Existing research has mainly utilised γH2AX to determine DNA damage caused by irradiation (Redon *et al.*, 2009; Lee *et al.*, 2019) or for early detection of carcinogenesis in rats (Toyoda and Ogawa, 2022; Toyoda *et al.*, 2023). As such, retroviral vector effect on DSB formation is not published on heavily, most such studies being used as markers to identify the infection by viruses such as herpesviridae (Tarakanova *et al.*, 2007; Botting, Lu and Triezenberg, 2015) rather than the effects of retroviral vectors utilised for gene therapy.

The vast information published on irradiation can also be used to develop a scale of DNA damage caused by retroviral vectors. The γH2AX assay work carried out in this thesis could be used to compare the DSB damage caused by radiation in order to demonstrate a scale of damage that is being caused by vectors. This information can highlight the damage IR radiation from the sun can cause to NHEJ impaired mechanisms. Thus patients for gene therapy who use the assay developed in this thesis to determine their susceptibility to DSB damage can also be advised on how much they should limit skin exposure to the sun.

Other causes for concern with the use of RV vectors are vector particles lacking a genome, referred to as empty vectors in this thesis. These empty vectors are known to

be included in a batch of LV or RV gene therapy vectors (Zhao *et al.*, 2008). At a basic level, these empty particles can be considered impurities, which can affect the efficacy and safety of the batch of vectors (Zhao *et al.*, 2008).

This thesis has shown in **chapter 3** that empty vectors lacking any genetic material cause DSBs comparable to those cause by complete vectors. Thus, this thesis has shown that it is necessary to develop methodologies to be able to clear empty vectors from batches of RV vectors to be used for gene therapy beyond the basic level of improving the efficacy of vectors.

Prior research has made use of low speed centrifugation to enhance the titres of retroviral vectors (Darling *et al.*, 2000). This method has also been used for much smaller particles, such as aplhaviruses (Rayaprolu *et al.*, 2020). The weight of empty vector particles might be low enough compared to complete vectors, to the point where low speed centrifugation could potentially be used to separate out empty vectors from a patch of retroviral vectors generated for gene therapy. Outside of this, research has made use of size-exclusion chromatography to determine the differences of empty viral vector particles from complete ones (Zhao *et al.*, 2008). This method could be used to determine more differences between the empty and complete vectors to allow for a method for the removal of the empty vectors.

A method through which retroviral vectors can cause unintended side effects is via genotoxicity. Genotoxicity is the property of a chemical agent to induce an uncontrolled change in the genetic material of a cell and leads to a mutagenic change in phenotype such as cancer (Ramezani, Hawley and Hawley, 2008).

Retroviral vectors can cause genotoxicity during the integration phase of their life cycle. As mentioned, the viral vector needs to induce the formation of DSBs in the host genome to allow the incorporation of the therapeutic gene into the host cell. This formation of DSBs does not occur at precise location, rather, the locations for the DSB formation can be semi-random within the open chromatin regions which are actively being transcribed (Albanese *et al.*, 2008; Baum *et al.*, 2004; Mitchell *et al.*, 2004; Wu *et al.*, 2003; Schröder *et al.*, 2002). This semi-random nature of the retroviral integration can cause integration to occur within a gene, or around a gene, leading to the activation, over-activation, under-activation or inactivation of a gene resulting in a genotoxic effect for the host (Nienhuis, Dunbar and Sorrentino, 2006). Genotoxic effects caused by retroviral vectors primarily involve oncogenesis (Modlich and Baum, 2009; Montini *et al.*,

2009; Cherepanov, 2007; Bushman *et al.*, 2005; Uren *et al.*, 2005; Baum *et al.*, 2004; Schröder *et al.*, 2002).

This thesis, in **chapter 4**, investigated the genotoxic effects of viral vectors along with empty viral vectors via the use of the *HPRT* gene knockout assay. It involves the detection of the knock-out of *HPRT* gene function, which is required in the salvage pathway for synthesising nucleotides (Themis *et al.*, 2003). The gene is present on the X chromosome (Zhang *et al.*, 1994), meaning that only one copy of the gene needs to be affected in male cells to determine gene knock out. This thesis showed that retroviral vectors cause *HPRT* knock at significant levels over the control as has been established in prior work (Themis *et al.*, 2003). However, this thesis has been able to test the effects of empty vectors on *HPRT* gene knock out and determined it was comparable to negative control levels. Since the complete vector caused higher amounts of genotoxicity than the empty vector, it is clear that a genome or genetic material is required to cause vector-mediated *HPRT* knockout; thus it is implicit that the viral vectors cause *HPRT* knockout via integration which the empty vectors are unable to do due to a lack of a genome or genetic material.

As we had seen in **chapter 3** that empty vectors cause a significant amount of DSBs, it is comparable to the DSB formation caused by complete vectors. Thus, it would seem, that DSB formation alone cannot cause gene knock out of the *HPRT* gene. However, this thesis has shown that when plasmids containing *gag-pol* are expressed to produce protein, these transfections cause the highest amount of genotoxicity. The proteins expressed by the *gag-pol* plasmids would include the integrase enzyme, which is responsible for DSB formation (Skalka and Katz, 2005). Since the *gag-pol* plasmid expressions lack a genome to integrate into the host cell, DSB formation alone did cause gene knock out of the *HPRT* gene.

It is quite possible that the *gag-pol* transfections produce the integrase protein at higher levels compared to the empty vectors due to the former allowing for pure expression of integrase protein, and the latter causing virus regulated integrase production. This would also lead to a higher amount of DSB formation which causes enough interruption in and/or around the *HPRT* gene that the error prone NHEJ mechanism is causing a build-up of many small errors, leading to a state of dysfunction in the *HPRT* gene.

Due to time constraints, only male cell lines with functional NHEJ mechanisms could be obtained for experimentation that could survive the primary selection step of the *HPRT* assay in HAT medium. For further work, if more male cell lines could be procured which

lacked NHEJ function, it could be determined how well they could undergo the *HPRT* assay. The results from the *gag-pol* transfections indicate that a very high amount of DSBs could even overcome the functional NHEJ mechanism of the cell lines that were utilised in this thesis, the MRC5 line and the V79 lines. Thus it would be interesting to note how the NHEJ impaired cell lines could fare with the *HPRT* assay when infected with the empty vectors. It could be that the more controlled amount of DSB formation caused by the empty vectors would be too much for the NHEJ impaired lines, also leading to genotoxicity.

If the above theory was true, that NHEJ impaired lines would undergo *HPRT* knockout, then the γH2AX assay mentioned earlier for gene therapy candidates could also reveal individuals who are more prone to genotoxicity following retroviral gene therapy; yet another piece of crucial information for such candidates.

Another possible cause for concern in the use of retroviral vectors is the presence of human endogenous retroviruses (HERVs) within the human genome. HERVs are viral genetic sequences present in the human genome which were derived from exogenous RV infections as the human genome has evolved. These sequences account for around 8% of the human genome (Feschotte and Gilbert, 2012; Dewannieux and Heidmann, 2013). Most HERVs are considered to lack infectious activity due to acquired disruption and silencing mutations rendering them non-functional over the course of evolution. However, HERVs may still be recognised as viral components by the human immune system and may lead to autoimmune disorders and cancers, though to what extent is not clear (Saini *et al.*, 2020; Dewannieux and Heidmann, 2013; Stoye, 2012).

Human cell lines used to package viral vectors for gene therapy can, at least at low levels, package these HERV sequences (Patience *et al.*, 2012). While thought to be non-infectious, their packaging into viral vectors does pose a risk to the efficacy and purity of vector batches.

This thesis wanted to explore the possibility of empty viral vectors formed in human producer cells to package the HERV sequences. These HERV particle could then reduce the efficacy of a viral vector batch, while also causing safety concerns by these vectors now being able to cause genotoxicity via insertional mutagenesis, much like complete retroviral vectors are able to.

These HERV vectors in a batch of vectors produced in human cells, could be integrated into highly transcribed areas in the genomes of patients and cause detrimental effects.

As mentioned, HERV sequences can cause autoimmune disorders such as immunosuppression (Grandi and Tramontano, 2018), can cause neurotoxic effects (Nair *et al.*, 2022) and have been seen to be involved in oncogenesis (Li *et* al., 2019; Agoni, Luha and Lenz. 2013; Büscher, K, 2005; Sauter *et al.*, 1995, Frank *et al.* 2008; Ishida *et al.*, 2008). Thus the incorporation of HERVs into empty vectors would pose a great risk to host cells should they be integrated into the genome at loci that are highly transcribed, or loci which are involved in gene regulation.

Experiments done for this thesis designed primers to amplify a known sequence of HERV, HERV K113, which has been seen in pieces and even in its entirety in multiple loci across the human genome (Wildschutte *et* al., 2016; van der Kuyl, 2012; Beimford *et al.*, 2008). HERV-K113 has been seen to be expressed in baculovirus expression vectors which were able to synthesise complete HERV-K113 particles (Boller *et al.*, 2007), making it a good candidate to investigate for incorporation into empty retroviral vectors due to known interaction with viral vectors. These primers would then be used to determine the presence of HERVs in non-human cells following infection by human producer cell line made retroviral vectors. The presence of HERVs in these non-human cells would indicate the integration of HERV-K113, thus showing that empty vectors could be incorporating HERV sequences which could go on to cause genotoxicity as well as other conditions.

The primers were successfully able to amplify a band of expected size which was targeted in the HERV-K113 sequence; although they also produced bands which were not of the expected size in smaller quantities. The presence of superfluous bands indicates that more than just the HERV K113 sequence in its entirety was being amplified. One reason the HERV K113 was not just amplified in its entirety is it being found in varying sequences ranging from its entire coding capacity to pieces of it at multiple loci in the human genome (Wildschutte *et* al., 2016; van der Kuyl, 2012; Beimford *et al.*, 2008). As shown in **figure 5.16**, the varying lengths and locations across the genome could explain the formation of some of the superfluous bands seen.

Evidence to suggest it was indeed a HERV that was amplified included the lack of any PCR product when the primers were used on non-human DNA such as mouse and Chinese hamster. This counts as evidence due to the fact that HERVs are evolutionarily tied to humans (Feschotte and Gilbert, 2012; Dewannieux and Heidmann, 2013), and thus could not be present in mouse or Chinese hamster DNA for the primers to amplify. These results also indicate that these particular cell lines are good candidates for the

next proposed part of the experiment to infect non-human cells with human produced retroviral vectors. If these cell lines were infected with human produced RV vectors, and the HERV primers are able to retrieve a PCR product from the transducted non-human cell line, it would then mean that the HERV would have most likely been incorporated into in the non-human cell lines by the human produced RV vectors.

More evidence to suggest that the targeted 7,500 bp sequence was also amplified was the expected restriction digest profile, indicating that the HERV sequence was present.

However, the presence of superfluous bands indicates that the primers are non-specific, and amplifying sequences that are not targeted. The restriction digest also had bands present at the 7,500 bp mark, indicating that either not all the HERV sequence was digested, or that there were other bands present at the 7,500 bp mark that were not the HERV sequence.

In order to better understand the products produced by the HERV 3 primer pair, a primer BLAST (Ye *et al.*, 2012), was conducted with the H3 primer pair on the human genome; the detailed results of which are shown in detail under **Appendix 5**. The primer BLAST showed that there were several superfluous products formed, with 20 of them being around 7,500 bp the general expected size of HERV sequence being targeted. Of these 20, 5 were the HERV K113 sequence at different loci. This shows that primer specificity could be improved to produce more specific products; although it has been shown that PCR experiments are poor at targeting large sequences within the human genome (Ye *et al.*, 2012). As such, it might be better to attempt to use other methods to find HERV sequence transfer to non-human cells via human produced retroviral vector transduction. However, even with such constraints, the expected band results from restriction digests still leaves this method open as an option.

A method to confirm that the band contained the correct HERV sequence involves sequencing the band for all the genetic sequences present (Dorit *et al.*, 2001). This would allow a confirmation the HERV sequence from the superfluous sequences.

Due to time constraints, experiments where non-human cells would be infected by empty vectors produced in human cells were not performed. These cells could then be analysed for the HERV sequence using the H3 PCR, and if present, would indicate the integration of HERV-K113 to the non-human cells via the empty vectors.

More HERV sequences which have been found in their entirety, such as HERV-Fs (Widegren *et al.*, 1996; Kjellman *et al.* 1999) and HERV-Hs (Jern *et al.*, 2005), could be

targeted to determine their integration via retroviral vectors being synthesised in human producer lines.

Other methods such as Southern blot tests and ELISA assays could be developed test the transfer of HERVs into non-human species via retroviral vectors. *In silico* methodologies, such as the use of RNAseq datasets with more sequence depth (more reads per sample), which have been used to locate HERV sequences (Hamann, Abida and Lange, 2023) could be used to locate more full length HERVs, which could then be analysed for genotoxicity.

In summary, this thesis has been able to show several of data which indicate that empty and complete retroviral vectors can cause damage to host cells via DSB formation and genotoxicity. A path is laid clear which shows the development of an assay to screen gene therapy candidates for NHEJ integrity which can assist these individuals with gene therapy options as well as lifestyle choices. It has also been shown that a possible simple PCR assay can be made to attempt to observe the transfer of HERV sequences from human producer cells to integration into non-human cells, the implications of which could reshape how retroviral gene therapy is considered in the future.

References

Abbaszadeh, F., Clingen, P.H., Arlett, C.F., Plowman, P.N., Bourton, E.C., Themis, M., Makarov, E.M., Newbold, R.F., Green, M.H.L. and Parris, C.N. (2010) 'A novel splice variant of the DNA-PKcs gene is associated with clinical and cellular radiosensitivity in a patient with xeroderma pigmentosum', *Journal of Medical Genetics,* 47(3), pp. 176-181. doi: 10.1136/jmg.2009.068866.

Addgene (2022a) *Addgene: pCMVR8.74.* Available at: https://www.addgene.org/22036/ (Accessed: 13/06/2022).

Addgene (2022b) *Addgene: pMD2.G.* Available at: https://www.addgene.org/12259/ (Accessed: 13/06/2022).

Agoni, L., Guha, C. and Lenz, J. (2013) 'Detection of Human Endogenous Retrovirus K (HERV-K) Transcripts in Human Prostate Cancer Cell Lines', *Frontiers in Oncology,* 3(1), pp. 180. doi: 10.3389/fonc.2013.00180.

Aiken, C. and Trono, D. (1995) 'Nef stimulates human immunodeficiency virus type 1 proviral DNA synthesis', *Journal of Virology,* 69(8), pp. 5048-5056. doi: 10.1128/JVI.69.8.5048-5056.1995.

Aiuti, A., Bachoud-Levi, A., Cattaneo, F., Roncarolo, M.G., Leen, A.M., Brenner, M.K., Mcneish, I.A., Lemoine, N.R., Chiocca, E.A., Waxman, D.J., Peschanski, M., BLESCH, A., Tuszynski, M.H., Strayer, D.S., Gao, G., Wilson, J.M., High, K.A. and Meneguzzi, G. (2007) 'Progress and Prospects: Gene Therapy Clinical Trials (Part 2)', *Gene Therapy,* 14(24), pp. 1555-1563. doi: 10.1038/sj.gt.3303033.

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. (2002) 'Immune reconstitution in ADA-SCID after PBL gene therapy and discontinuation of enzyme replacement', *Nature Medicine,* 8(5), pp. 423-425. doi: 10.1038/nm0502-423.

Akagi, K., Suzuki, T., Stephens, R.M., Jenkins, N.A. and Copeland, N.G. (2004) 'RTCGD: retroviral tagged cancer gene database', *Nucleic Acids Research,* 32(suppl-1), pp. D523-D527. doi: 10.1093/nar/gkh013.

Åkerlund, E., Cappellini, F., Di Bucchianico, S., Islam, S., Skoglund, S., Derr, R., Odnevall Wallinder, I., Hendriks, G., Karlsson, H.L. and Johnson, G. (2018) 'Genotoxic and mutagenic properties of Ni and NiO nanoparticles investigated by comet assay, γ-H2AX staining, Hprt mutation assay and ToxTracker reporter cell lines', *Environmental and Molecular Mutagenesis,* 59(3), pp. 211-222. doi: 10.1002/em.22163.

Akkina, R.K., Walton, R.M., Chen, M.L., Li, Q.X., Planelles, V. and Chen, I.S. (1996) 'High-efficiency gene transfer into CD34+ cells with a human immunodeficiency virus type 1-based retroviral vector pseudotyped with vesicular stomatitis virus envelope glycoprotein G', *Journal of Virology,* 70(4), pp. 2581-2585. doi: 10.1128/JVI.70.4.2581-2585.1996.

Alam, T., Wai, P., Held, D., Vakili, S.T.T., Forsberg, E. and Sollinger, H. (2013) 'Correction of Diabetic Hyperglycemia and Amelioration of Metabolic Anomalies by Minicircle DNA Mediated Glucose-Dependent Hepatic Insulin Production', *PLoS ONE,* 8(6), pp. e67515. doi: 10.1371/journal.pone.0067515.

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,* 3(6), pp. e2413. doi: 10.1371/journal.pone.0002413.

Andersen, J.L. and Planelles, V. (2005) 'The Role of Vpr in HIV-1 Pathogenesis', *Current HIV Research,* 3(1), pp. 43-51. doi: 10.2174/1570162052772988.

Andtbacka, R.H.I., Kaufman, H.L., Collichio, F., Amatruda, T., Senzer, N., Chesney, J., Delman, K.A., Spitler, L.E., Puzanov, I., Agarwala, S.S., Milhem, M., Cranmer, L., Curti, B., Lewis, K., Ross, M., Guthrie, T., Linette, G.P., Daniels, G.A., Harrington, K., Middleton, M.R., Miller, J., Wilson H, Zager, J.S., Ye, Y., Yao, B., Li, A., Doleman, S., VanderWalde, A., Gansert, J. and Coffin, R.S. (2015) 'Talimogene Laherparepvec Improves Durable Response Rate in Patients With Advanced Melanoma', *Journal of Clinical Oncology,* 33(25), pp. 2780-2788. doi: 10.1200/JCO.2014.58.3377.

Arlett, C.F., Green, M.H.L., Priestley, A., Harcourt, S.A. and Mayne, L.V. (1988) 'Comparative Human Cellular Radiosensitivity: I. The Effect of SV40 Transformation and Immortalisation on the Gamma-irradiation Survival of Skin Derived Fibroblasts from Normal Individuals and from Ataxia-telangiectasia Patients and Heterozygotes', *International Journal of Radiation Biology,* 54(6), pp. 911-928. doi: 10.1080/09553008814552321.

Arrigo, S.J., Weitsman, S., Rosenblatt, J.D. and Chen, I. (1989) 'Analysis of rev gene function on human immunodeficiency virus type 1 replication in lymphoid cells by using a quantitative polymerase chain reaction method', *Journal of Virology,* 63(11), pp. 4878-4881.

Baier, K., Morell, M. and O'Carroll, I.P. (2022) 'Mutational analysis of the oligomerization domain of the HERV-K Rec protein', *The FASEB Journal,* 36(Suppl 1), pp. n/a. doi: 10.1096/fasebj.2022.36.S1.R4479.

Barber, M.A. (1911) 'A Technic for the Inoculation of Bacteria and Other Substances Into Living Cells', *The Journal of Infectious Diseases,* 8(3), pp. 348-360. doi: 10.1093/infdis/8.3.348.

Baum, C., Itoh, K., Meyer, J., Laker, C., Ito, Y. and Ostertag, W. (1997) 'The potent enhancer activity of the polycythemic strain of spleen focus-forming virus in hematopoietic cells is governed by a binding site for Sp1 in the upstream control region and by a unique enhancer core motif, creating an exclusive target for PEBP/CBF', *Journal of Virology,* 71(9), pp. 6323-6331. doi: 10.1128/JVI.71.9.6323-6331.1997.

Baum, C., Hunt, N., Hildinger, M., Eckert, H., Zaehres, H., Richters, A., John, J., Löhler, J. and Ostertag, W. (1998) 'Cis-Active Elements of Friend Spleen Focus-Forming Virus: From Disease Induction to Disease Prevention', *Acta Haematologica,* 99(3), pp. 156-164. doi: 10.1159/000040830.

Baum, C., Schambach, A., Bohne, J. and Galla, M. (2006) 'Retrovirus Vectors: Toward the Plentivirus?', *Molecular Therapy,* 13(6), pp. 1050-1063. doi: 10.1016/j.ymthe.2006.03.007.

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,* 9(1), pp. 5-13. doi: 10.1016/j.ymthe.2003.10.013.

Beimforde, N., Hanke, K., Ammar, I., Kurth, R. and Bannert, N. (2007) 'Molecular cloning and functional characterization of the human endogenous retrovirus K113', *Virology,* 371(1), pp. 216-225. doi: 10.1016/j.virol.2007.09.036.

Biasco, L., Rothe, M., Büning, H. and Schambach, A. (2018) 'Analyzing the Genotoxicity of Retroviral Vectors in Hematopoietic Cell Gene Therapy', *Molecular Therapy. Methods & Clinical Development,* 8(C), pp. 21-30. doi: 10.1016/j.omtm.2017.10.002.

Björgvinsdottir, H., Ding, C., Pech, N., Gifford, M.A., Ling Lin Li and Dinauer, M.C. (1997) 'Retroviral-mediated gene transfer of gp91phox into bone marrow cells rescues defect in host defense against Aspergillus fumigatus in murine X-linked chronic granulomatous disease', *Blood,* 89(1), pp. 41-48. doi: 10.1182/blood.V89.1.41.41_41_48.

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,* 270(5235), pp. 475-480. doi: 10.1126/science.270.5235.475.

Bolhassani, A., Khavari, A. and Orafa, Z. (2014) 'Electroporation – Advantages and Drawbacks for Delivery of Drug, Gene and Vaccine', in Sezer, A.D. (ed.) *Application of Nanotechnology in Drug Delivery*. 1st edn. London: IntechOpen, pp. 369-397.

Boller, K., Schonfeld, K., Lischer, S., Fischer, N., Hoffmann, A., Kurth, R. and Tonjes, R.R. (2008) 'Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles', *Journal of General Virology,* 89(2), pp. 567-572. doi: 10.1099/vir.0.83534-0.

Botting, C., Lu, X. and Triezenberg, S.J. (2016) 'H2AX Phosphorylation and DNA Damage Kinase Activity Are Dispensable for Herpes Simplex Virus Replication', *Virology Journal,* 13(15), pp. 15. doi: 10.1186/s12985-016-0470-1.

Borenfreund, E. and Bendich, A. (1961) 'A Study of the Penetration of Mammalian Cells by Deoxyribonucleic Acids', *Journal of Biophysical and Biochemical Cytology,* 9(1), pp. 81-91.

Bosch, M.L., Earl, P.L., Fargnoli, K., Picciafuoco, S., Giombini, F., Wong-Staal, F. and Franchini, G. (1989) 'Identification of the fusion peptide of primate immunodeficiency viruses', *Science,* 244(4905), pp. 694-697. doi: 10.1126/science.2541505.

Bradley, M.O., Bhuyan, B., Francis, M.C., Langenbach, R., Peterson, A. and Huberman, E. (1981) 'Mutagenesis by chemical agents in V79 Chinese hamster cells: A review and analysis of the literature: A report of the gene-tox program', *Mutation research,* 87(2), pp. 81-142. doi: 10.1016/0165-1110(81)90029-4.

Brandsma, I. and Gent, D.C. (2012) 'Pathway choice in DNA double strand break repair: observations of a balancing act', *Genome Integrity,* 3(1), pp. 9. doi: 10.1186/2041-9414-3-9.

Briggs, J.A.G., Wilk, T., Welker, R., Kräusslich, H. and Fuller, S.D. (2003) 'Structural organization of authentic, mature HIV-1 virions and cores', *The EMBO Journal,* 22(7), pp. 1707-1715. doi: 10.1093/emboj/cdg143.

Brun, S., Faucon-Biguet, N. and Mallet, J. (2003) 'Optimization of transgene expression at the posttranscriptional level in neural cells: implications for gene therapy', *Molecular Therapy,* 7(6), pp. 782-789. doi: 10.1016/S1525-0016(03)00097-2.

Buchholz, C.J., Mühlebach, M.D. and Cichutek, K. (2009) 'Lentiviral vectors with measles virus glycoproteins – dream team for gene transfer?', *Trends in Biotechnology (Regular Ed.),* 27(5), pp. 259-265. doi: 10.1016/j.tibtech.2009.02.002.

Bukrinsky, M.I., Sharova, N., Mcdonald, T.L., Pushkarskaya, T., Tarpley, W.G. and Stevenson, M. (1993) 'Association of Integrase, Matrix, and Reverse Transcriptase Antigens of Human Immunodeficiency Virus Type 1 with Viral Nucleic Acids Following Acute Infection', *Proceedings of the National Academy of Sciences - PNAS,* 90(13), pp. 6125-6129. doi: 10.1073/pnas.90.13.6125.

Burns, J.C., Friedmann, T., Driever, W., Burrascano, M. and Jiing-Kuan Yee (1993) 'Vesicular Stomatitis Virus G Glycoprotein Pseudotyped Retroviral Vectors: Concentration to Very High

Titer and Efficient Gene Transfer into Mammalian and Nonmammalian Cells', *Proceedings of the National Academy of Sciences - PNAS,* 90(17), pp. 8033-8037. doi: 10.1073/pnas.90.17.8033.

Büscher, K., Trefzer, U., Hofmann, M., Sterry, W., Kurth, R. and Denner, J. (2005) 'Expression of human endogenous retrovirus K in melanomas and melanoma cell lines', *Cancer Research,* 65(10), pp. 4172-4180. doi: 10.1158/0008-5472.CAN-04-2983.

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,* 3(11), pp. 848-858. doi: 10.1038/nrmicro1263.

Byrnes, A.P., Rusby, J.E., Wood, M.J.A. and Charlton, H.M. (1995) 'Adenovirus gene transfer causes inflammation in the brain', *Neuroscience,* 66(4), pp. 1015-1024. doi: 10.1016/0306-4522(95)00068-T.

Cara, A. and Reitz, M.S. (1997) 'New insight on the role of extrachromosomal retroviral DNA', *Leukemia,* 11(9), pp. 1395-1399. doi: 10.1038/sj.leu.2400776.

Cartier, N., Hacein-Bey-Abina, S., Bartholomae, C.C., Veres, G., Schmidt, M., Kutschera, I., Vidaud, M., Abel, U., Dal-Cortivo, L., Caccavelli, L., Mahlaoui, N., Kiermer, V., Mittelstaedt, D., Bellesme, C., Lahlou, N., Lefrere, F., Blanche, S., Audit, M., Payen, E., Leboulch, P., l'Homme, B., Bougneres, P., Von Kalle, C., Fischer, A., Cavazzana-Calvo, M. and Aubourg, P. (2009) 'Hematopoietic Stem Cell Gene Therapy with a Lentiviral Vector in X-Linked Adrenoleukodystrophy', *Science,* 326(5954), pp. 818-823. doi: 10.1126/science.1171242.

Cavazzana-Calvo, M. and Fischer, A. (2007) 'Gene therapy for severe combined immunodeficiency: are we there yet?', *The Journal of Clinical Investigation,* 117(6), pp. 1456-1465. doi: 10.1172/JCI30953.

Cavazzana-Calvo, M., Hacein-Bey, S., Basile, G.d.S., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J., Bousso, P., Deist, F.L. and Fischer, A. (2000) 'Gene Therapy of Human Severe Combined Immunodeficiency (SCID)-X1 Disease', *Science,* 288(5466), pp. 669-672. doi: 10.1126/science.288.5466.669.

Cavazzana-Calvo, M. and Hacein-Bey-Abina, S. (2001) 'Correction of genetic blood defects by gene transfer', *Current Opinion in Hematology,* 8(6), pp. 360-367. doi: 10.1097/00062752-200111000-00008.

Cavazzana-Calvo, M., Payen, E., Cavallesco, R., Gillet-Legrand, B., Caccavelli, L., Sgarra, R., Maouche-Chretien, L., Bernaudin, F., Giroe, R., Dorazio, R., Mulder, G., Polack, A., Negre, O., Bank, A., Soulier, J., Larghero, J., Kabbara, N., Dalle, B., Gourmel, B., Socie, G., Chretien, S., Cartier, N., Aubourg, P., WANG, G., Fischer, A., Cornetta, K., Galacteros, F., Beuzard, Y., Gluckman, E., Bushman, F., Hacein-Bey-Abina, S., Leboulch, P., Hehir, K., Fusil, F., Down, J.,

Cesana, D., Ranzani, M., Volpin, M., Bartholomae, C., Duros, C., Artus, A., Merella, S., Benedicenti, F., Sergi Sergi, L., Sanvito, F., Brombin, C., Nonis, A., Serio, C.D., Doglioni, C., von Kalle, C., Schmidt, M., Cohen-Haguenauer, O., Naldini, L. and Montini, E. (2014) 'Uncovering and Dissecting the Genotoxicity of Self-inactivating Lentiviral Vectors In Vivo', *Molecular Therapy,* 22(4), pp. 774-785. doi: 10.1038/mt.2014.3.

Ci, Y., Yang, Y., Xu, C. and Shi, L. (2018) 'Vesicular Stomatitis Virus G Protein Transmembrane Region Is Crucial for The Hemi-Fusion to Full Fusion Transition', *Scientific Reports,* 8(1), pp. 10669-11. doi: 10.1038/s41598-018-28868-y.

Denaro, M., Brady, T. and Westerman, K. (2010) 'Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia', *Nature,* 467(7313), pp. 318-322. doi: 10.1038/nature09328.

Cervera, L., Gòdia, F., Tarrés-Freixas, F., Aguilar-Gurrieri, C., Carrillo, J., Blanco, J. and Gutiérrez-Granados, S. (2019) 'Production of HIV-1-based virus-like particles for vaccination: achievements and limits', *Applied Microbiology and Biotechnology,* 103(18), pp. 7367-7384. doi: 10.1007/s00253-019-10038-3.

Chan, D.C. and Kim, P.S. (1998) 'HIV entry and its inhibition', *Cell,* 93(5), pp. 681-684. doi: 10.1016/s0092-8674(00)81430-0.

Chang, L.J., Liu, X. and He, J. (2005) 'Lentiviral siRNAs targeting multiple highly conserved RNA sequences of human immunodeficiency virus type 1', *Gene Therapy,* 12(14), pp. 1133-1144. doi: 10.1038/sj.gt.3302509.

Charneau, P., Mirambeau, G., Roux, P., Paulous, S., Buc, H. and Clavel, F. (1994) 'HIV-1 Reverse Transcription A Termination Step at the Center of the Genome', *Journal of Molecular Biology,* 241(5), pp. 651-662. doi: 10.1006/jmbi.1994.1542.

Check, E. (2002) 'A tragic setback', *Nature,* 420(6912), pp. 116-118. doi: 10.1038/420116a.

Cherepanov, P. (2007) 'LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro', *Nucleic Acids Research,* 35(1), pp. 113-124. doi: 10.1093/nar/gkl885.

Chesnoy, S. and Huang, L. (2000) 'Structure and Function of Lipid-DNA Complexes for Gene Delivery', *Annual Review of Biophysics and Biomolecular Structure,* 29(1), pp. 27-47. doi: 10.1146/annurev.biophys.29.1.27.

Chirmule, N., Propert, K., Magosin, S., Qian, Y., Qian, R. and Wilson, J. (1999) 'Immune responses to adenovirus and adeno-associated virus in humans', *Gene Therapy,* 6(9), pp. 1574-1583. doi: 10.1038/sj.gt.3300994.

Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P.D., Wu, L., Mackay, C.R., LaRosa, G., Newman, W., Gerard, N., Gerard, C. and Sodroski, J. (1996) 'The β-Chemokine Receptors CCR3 and CCR5 Facilitate Infection by Primary HIV-1 Isolates', *Cell,* 85(7), pp. 1135-1148. doi: 10.1016/S0092-8674(00)81313-6.

Chung, H., Lee, H., Kim, H. and Hong, S. (2011) 'The Mechanical Agitation Method of Gene Transfer for Ex-Vivo Gene Therapy', in Yuan, X. (ed.) *Non-Viral Gene Therapy*. 1st edn. London: InTech, pp. 91-104.

Coffin, J.M. (1996) 'Retroviridae: The viruses and their replication', in Fields, B.N., Knipe, D.M. and Howley, P.M. (eds.) *Fundamental Virology*. 3rd edn. Philadelphia: Lippincott-Raven Publishers, pp. 763-843.

Cohen, C.J., Lock, W.M. and Mager, D.L. (2009) 'Endogenous retroviral LTRs as promoters for human genes: A critical assessment', *Gene,* 448(2), pp. 105-114. doi: 10.1016/j.gene.2009.06.020.

Coil, D.A. and Miller, A.D. (2004) 'Phosphatidylserine Is Not the Cell Surface Receptor for Vesicular Stomatitis Virus', *Journal of Virology,* 78(20), pp. 10920-10926. doi: 10.1128/JVI.78.20.10920-10926.2004.

Collas, P. and Aleström, P. (1996) 'Nuclear localization signal of SV40 T antigen directs import of plasmid DNA into sea urchin male pronuclei in vitro', *Molecular Reproduction and Development,* 45(4), pp. 431-438. doi: 10.1002/(SICI)1098-2795(199612)45:4<431::AID-MRD4>3.0.CO;2-S.

Collins, K.L., Chen, B.K., Kalams, S.A., Walker, B.D. and Baltimore, D. (1998) 'HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes', *Nature,* 391(6665), pp. 397-401. doi: 10.1038/34929.

Contreras-Galindo, R., Kaplan, M.H., Dube, D., Gonzalez-Hernandez, M.J., Chan, S., Meng, F., Dai, M., Omenn, G.S., Gitlin, S.D. and Markovitz, D.M. (2015) 'Human Endogenous Retrovirus Type K (HERV-K) Particles Package and Transmit HERV-K–Related Sequences', *Journal of Virology,* 89(14), pp. 7187-7201. doi: 10.1128/JVI.00544-15.

Contreras-Galindo, R., Kaplan, M.H., Markovitz, D.M., Lorenzo, E. and Yamamura, Y. (2006) 'Detection of HERV-K(HML-2) Viral RNA in Plasma of HIV Type 1-Infected Individuals', *AIDS research and human retroviruses,* 22(10), pp. 979-984. doi: 10.1089/aid.2006.22.979.

Contreras-Galindo, R., Lopez, P., Velez, R. and Yamamura, Y. (2007) 'HIV-1 Infection Increases the Expression of Human Endogenous Retroviruses Type K (HERV-K) in Vitro', *AIDS Research and Human Retroviruses,* 23(1), pp. 116-122. doi: 10.1089/aid.2006.0117.

Cornu, T.I., Mussolino, C., Müller, M.C., Wehr, C., Kern, W.V. and Cathomen, T. (2021) 'HIV Gene Therapy: An Update', *Human Gene Therapy,* 32(1-2), pp. 52-65. doi: 10.1089/hum.2020.159.

Cosset, F.L., Takeuchi, Y., Battini, J.L., Weiss, R.A. and Collins, M.K. (1995) 'High-titer packaging cells producing recombinant retroviruses resistant to human serum', *Journal of Virology,* 69(12), pp. 7430-7436. doi: 10.1128/JVI.69.12.7430-7436.1995.

Coutelle, C., Themis, M., Waddington, S.N., Buckley, S.M.K., Gregory, L.G., Nivsarkar, M.S., David, A.L., Peebles, D., Weisz, B. and Rodeck, C. (2005) 'Gene Therapy Progress and Prospects: Fetal gene therapy - first proofs of concept - some adverse effects', *Gene Therapy,* 12(22), pp. 1601-1607. doi: 10.1038/sj.gt.3302632.

Craigie, R. (2001) 'HIV Integrase, a Brief Overview from Chemistry to Therapeutics', *Journal of Biological Chemistry,* 276(26), pp. 23213-23216. doi: 10.1074/jbc.R100027200.

Crick, F. (1970) 'Central Dogma of Molecular Biology', *Nature (London),* 227(5258), pp. 561-563. doi: 10.1038/227561a0.

Crystal, R.G. and O'Connor, T.P. (2006) 'Genetic medicines: treatment strategies for hereditary disorders', *Nature Reviews Genetics,* 7(4), pp. 261-276. doi: 10.1038/nrg1829.

Daniel, R., Katz, R.A., Merkel, G., Hittle, J.C., Yen, T.J. and Skalka, A.M. (2001) 'Wortmannin Potentiates Integrase-Mediated Killing of Lymphocytes and Reduces the Efficiency of Stable Transduction by Retroviruses', *Molecular and Cellular Biology,* 21(4), pp. 1164-1172. doi: 10.1128/MCB.21.4.1164-1172.2001.

Daniel, R., Greger, J.G., Katz, R.A., Taganov, K.D., Xiaoyun Wu, Kappes, J.C. and Skalka, A.M. (2004) 'Evidence that stable retroviral transduction and cell survival following DNA integration depend on components of the nonhomologous end joining repair pathway', *Journal of Virology,* 78(16), pp. 8573-8581. doi: 10.1128/JVI.78.16.8573-8581.2004.

Daniel, R., Katz, R.A. and Skalka, A.M. (1999) 'A role for DNA-PK in retroviral DNA integration', *Science,* 284(5414), pp. 644-647. doi: 10.1126/science.284.5414.644.

Darling, D., Hughes, C., Galea-Lauri, J., Gäken, J., Trayner, I.D., Kuiper, M. and Farzaneh, F. (2000) 'Low-speed centrifugation of retroviral vectors absorbed to a particulate substrate: a highly effective means of enhancing retroviral titre', *Gene Therapy,* 7(11), pp. 914-923. doi: 10.1038/sj.gt.3301201.

Das, S.R. and Jameel, S. (2005) 'Biology of the HIV Nef protein', *Indian Journal of Medical Research,* 121(4), pp. 315-332. Obtained from: https://ijmr.icmr.org.in/ijmr/index.aspx

Daud, A.I., Deconti, R.C., Heller, R., Andrews, S., Urbas, P., Riker, A.I., Sondak, V.K., Munster, P.N., Sullivan, D.M., Ugen, K.E. and Messina, J.L. (2008) 'Phase I Trial of Interleukin-12 Plasmid Electroporation in Patients With Metastatic Melanoma', *Journal of Clinical Oncology,* 26(36), pp. 5896-5903. doi: 10.1200/JCO.2007.15.6794.

Davis, H.L., Whalen, R.G. and Demeneix, B.A. (1993) 'Direct gene transfer into skeletal muscle in vivo: factors affecting efficiency of transfer and stability of expression', *Human Gene Therapy,* 4(2), pp. 151-159. doi: 10.1089/hum.1993.4.2-151.

Demaison, C., Parsley, K., Brouns, G., Scherr, M., Battmer, K., Kinnon, C., Grez, M. and Thrasher, A.J. (2002a) 'High-level transduction and gene expression in hematopoietic repopulating cells using a human imunodeficiency virus type 1-based lentiviral vector containing an internal spleen focus forming virus promoter', *Human Gene Therapy,* 13(7), pp. 803-813. doi: 10.1089/10430340252898984.

Demaison, C., Parsley, K., Brouns, G., Scherr, M., Battmer, K., Kinnon, C., Grez, M. and Thrasher, A.J. (2002b) 'High-level transduction and gene expression in hematopoietic repopulating cells using a human immunodeficiency [correction of imunodeficiency] virus type 1-based lentiviral vector containing an internal spleen focus forming virus promoter', *Human Gene Therapy,* 13(7), pp. 803-813. doi: 10.1089/10430340252898984.

Dewannieux, M. and Heidmann, T. (2013) 'Endogenous retroviruses: acquisition, amplification and taming of genome invaders', *Current Opinion in Virology,* 3(6), pp. 646-656. doi: 10.1016/j.coviro.2013.08.005.

Dolei, A., Ibba, G., Piu, C. and Serra, C. (2019) 'Expression of HERV Genes as Possible Biomarker and Target in Neurodegenerative Diseases', *International Journal of Molecular Sciences,* 20(15), pp. 3706. doi: 10.3390/ijms20153706.

Donello, J.E., Loeb, J.E. and Hope, T.J. (1998) 'Woodchuck Hepatitis Virus Contains a Tripartite Posttranscriptional Regulatory Element', *Journal of Virology,* 72(6), pp. 5085-5092. doi: 10.1128/JVI.72.6.5085-5092.1998.

Dorit, R.L., Ohara, O., Hwang, C.B., Kim, J.B. and Blackshaw, S. (2001) 'Direct DNA Sequencing of PCR Products', *Current Protocols in Molecular Biology,* Chapter 15(1), pp. Unit 15.2. doi: 10.1002/0471142727.mb1502s56.

Dragic, T., Litwin, V., Allaway, G.P., Martin, S.R., Huang, Y., Nagashima, K.A., Cayanan, C., Maddon, P.J., Koup, R.A., Moore, J.P. and Paxton, W.A. (1996) 'HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5', *Nature,* 381(6584), pp. 667-673. doi: 10.1038/381667a0.

Du, X., Wang, J., Zhou, Q., Zhang, L., Wang, S., Zhang, Z. and Yao, C. (2018) 'Advanced Physical Techniques for Gene Delivery Based on Membrane Perforation', *Drug Delivery,* 25(1), pp. 1516-1525. doi: 10.1080/10717544.2018.1480674.

Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D. and Naldini, L. (1998) 'A Third-Generation Lentivirus Vector with a Conditional Packaging System', *Journal of Virology,* 72(11), pp. 8463-8471. doi: 10.1128/JVI.72.11.8463-8471.1998.

Durand, S. and Cimarelli, A. (2011) 'The inside out of lentiviral vectors', *Viruses,* 3(2), pp. 132-159. doi: 10.3390/v3020132.

Eglitis, M.A., Kantoff, P., Gilboa, E. and Anderson, W.F. (1985) 'Gene expression in mice after high efficiency retroviral-mediated gene transfer', *Science,* 230(4732), pp. 1395-1398. doi: 10.1126/science.2999985.

Eisenman, R.N. and Vogt, V.M. (1978) 'The biosynthesis of oncovirus proteins', *Biochimica et Biophysica Acta. Reviews on Cancer,* 473(3), pp. 187-239. doi: 10.1016/0304-419X(78)90014-8.

Engelman, A., Mizuuchi, K. and Craigie, R. (1991) 'HIV-1 DNA integration: Mechanism of viral DNA cleavage and DNA strand transfer', *Cell,* 67(6), pp. 1211-1221. doi: 10.1016/0092-8674(91)90297-C.

Escarpe, P., Zayek, N., Chin, P., Borellini, F., Zufferey, R., Veres, G. and Kiermer, V. (2003) 'Development of a sensitive assay for detection of replication-competent recombinant lentivirus in large-scale HIV-based vector preparations', *Molecular Therapy,* 8(2), pp. 332-341. doi: 10.1016/S1525-0016(03)00167-9.

Everson, E.M., Olzsko, M.E., Leap, D.J., Hocum, J.D. and Trobridge, G.D. (2016) 'A Comparison of Foamy and Lentiviral Vector Genotoxicity in Scid-Repopulating Cells Shows Foamy Vectors Are Less Prone to Clonal Dominance', *Molecular Therapy: Methods & Clinical Development,* 3(C), pp. 16048. doi: 10.1038/mtm.2016.48.

Fehse, B. and Roeder, I. (2008) 'Insertional mutagenesis and clonal dominance: biological and statistical considerations', *Gene Therapy,* 15(2), pp. 143-153. doi: 10.1038/sj.gt.3303052.

Feschotte, C. and Gilbert, C. (2012) 'Endogenous viruses: insights into viral evolution and impact on host biology', *Nature Reviews Genetics,* 13(4), pp. 283-296. doi: 10.1038/nrg3199.

Finston, W.I. and Champoux, J.J. (1984) 'RNA-primed initiation of Moloney murine leukemia virus plus strands by reverse transcriptase in vitro', *Journal of Virology,* 51(1), pp. 26-33. doi: 10.1128/JVI.51.1.26-33.1984.

Fischer, A. and Cavazzana-Calvo, M. (2008) 'Gene therapy of inherited diseases', *The Lancet,* 371(9629), pp. 2044-2047.

Ford, D.K. and Yerganian, G. (1958) 'Observations on the chromosomes of Chinese hamster cells in tissue culture', *Journal of the National Cancer Institute,* 21(2), pp. 393-425. doi: 10.1093/jnci/21.2.393

Fouchier, R.A., Simon, J.H., Jaffe, A.B. and Malim, M.H. (1996) 'Human immunodeficiency virus type 1 Vif does not influence expression or virion incorporation of gag-, pol-, and env-encoded proteins', *Journal of Virology,* 70(12), pp. 8263-8269. doi: 10.1128/JVI.70.12.8263-8269.1996.

Frank, O., Verbeke, C., Schwarz, N., Mayer, J., Fabarius, A., Hehlmann, R., Leib-Mösch, C. and Seifarth, W. (2008) 'Variable Transcriptional Activity of Endogenous Retroviruses in Human Breast Cancer', *Journal of Virology,* 82(4), pp. 1808-1818. doi: 10.1128/JVI.02115-07.

Frecha, C., Szécsi, J., Cosset, F. and Verhoeyen, E. (2008) 'Strategies for Targeting Lentiviral Vectors', *Current Gene Therapy,* 8(6), pp. 449-460. doi: 10.2174/156652308786848003.

Fung, H. and Gerson, S.L. (2014) 'Chapter 3 - Viral Insertion Site Detection and Analysis in Cancer Gene Therapy', in Lattime, E.C. and Gerson, S.L. (eds.) *Gene Therapy of Cancer.* 3rd edn. San Diego: Academic Press, pp. 35-46.

Gadek, K.E., Wang, H., Hall, M.N., Sungello, M., Libby, A., MacLaskey, D., Eckel, R.H. and Olwin, B.B. (2018) 'Striated muscle gene therapy for the treatment of lipoprotein lipase deficiency', *PLoS ONE,* 13(1), pp. e0190963. doi: 10.1371/journal.pone.0190963.

Gaspar, H.B., Parsley, K.L., Howe, S., King, D., Gilmour, K.C., Sinclair, J., Brouns, G., Schmidt, M., Von Kalle, C., Barington, T., Jakobsen, M.A., Christensen, H.O., Al Ghonaium, A., White, H.N., Smith, J.L., Levinsky, R.J., Ali, R.R., Kinnon, C. and Thrasher, A.J. (2004) 'Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector', *The Lancet,* 364(9452), pp. 2181-2187. doi: 10.1016/S0140-6736(04)17590-9.

Gifford, R.J., Blomberg, J., Coffin, J.M., Fan, H., Heidmann, T., Mayer, J., Stoye, J., Tristem, M. and Johnson, W.E. (2018) 'Nomenclature for endogenous retrovirus (ERV) loci', *Retrovirology,* 15(1), pp. 59. doi: 10.1186/s12977-018-0442-1.

Gillet, J., Macadangdang, B., Fathke, R.L., Gottesman, M.M. and Kimchi-Sarfaty, C. (2009) 'The Development of Gene Therapy: From Monogenic Recessive Disorders to Complex Diseases Such as Cancer'*Gene Therapy of Cancer* Totowa, NJ: Humana Press, pp. 5-54.

Ginn, S.L., Amaya, A.K., Alexander, I.E., Edelstein, M. and Abedi, M.R. (2018) 'Gene therapy clinical trials worldwide to 2017: An update', *Journal of Gene Medicine,* 20(5), pp. e3015-n/a. doi: 10.1002/jgm.3015.

Glover, D.J., Lipps, H.J. and Jans, D.A. (2005) 'Towards safe, non-viral therapeutic gene expression in humans', *Nature Reviews Genetics,* 6(4), pp. 299-310. doi: 10.1038/nrg1577.

González, M.E. (2015) 'Vpu Protein: The Viroporin Encoded by HIV-1', *Viruses,* 7(8), pp. 4352-4368. doi: 10.3390/v7082824.

Gorman, L., Suter, D., Emerick, V., Schümperli, D. and Kole, R. (1998) 'Stable Alteration of Pre-mRNA Splicing Patterns by Modified U7 Small Nuclear RNAs', *Proceedings of the National Academy of Sciences - PNAS,* 95(9), pp. 4929-4934. doi: 10.1073/pnas.95.9.4929.

Gouvarchin, G.H.E., Bolandian, M., Dorostkar, R., Jafari, A. and Pour, M.F. (2020) 'Concise review on optimized methods in production and transduction of lentiviral vectors in order to facilitate immunotherapy and gene therapy', *Biomedicine & Pharmacotherapy,* 128, pp. 110276. doi: 10.1016/j.biopha.2020.110276.

Grandi, N. and Tramontano, E. (2018) 'Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses', *Frontiers in Immunology,* 9(1), pp. 2039. doi: 10.3389/fimmu.2018.02039.

Grosovsky, A.J., Skandalis, A., Hasegawa, L. and Walter, B.N. (1993) 'Insertional inactivation of the tk locus in a human B lymphoblastoid cell line by a retroviral shuttle vector', *Mutation research,* 289(2), pp. 297-308.

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. and Cavazzana-Calvo, M. (2008) 'Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1', *The Journal of Clinical Investigation,* 118(9), pp. 3132-3142. doi: 10.1172/JCI35700.

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. and Cavazzana-Calvo, M. (2010) 'Efficacy of Gene Therapy for X-Linked Severe Combined Immunodeficiency', *The New England Journal of Medicine,* 363(4), pp. 355-364. doi: 10.1056/NEJMoa1000164.

Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J., Fraser, C.C., Cavazzana-Calvo, M. and Fischer, A. (2003) 'A Serious Adverse Event after Successful Gene Therapy for X-Linked Severe Combined Immunodeficiency', *The New England Journal of Medicine,* 348(3), pp. 255-256. doi: 10.1056/NEJM200301163480314.

Haddad-Mashadrizeh, A., Zomorodipour, A., Izadpanah, M., Sam, M.R., Ataei, F., Sabouni, F. and Hosseini, S.J. (2009) 'A systematic study of the function of the human β-globin introns on the

expression of the human coagulation factor IX in cultured Chinese hamster ovary cells', *Journal of Gene Medicine,* 11(10), pp. 941-950. doi: 10.1002/jgm.1367.

Halwani, R., Shan Cen, Javanbakht, H., Saadatmand, J., Sunghoon Kim, Shiba, K. and Kleiman, L. (2004) 'Cellular Distribution of Lysyl-tRNA Synthetase and Its Interaction with Gag during Human Immunodeficiency Virus Type 1 Assembly', *Journal of Virology,* 78(14), pp. 7553-7564. doi: 10.1128/JVI.78.14.7553-7564.2004.

Hamann, M.V., Adiba, M. and Lange, U.C. (2023) 'Confounding Factors in Profiling of Locus-Specific Human Endogenous Retrovirus (HERV) Transcript Signatures in Primary T Cells Using Multi-Study-Derived Datasets', *BMC Medical Genomics,* 16(1), pp. 68. doi: 10.1186/s12920-023-01486-y.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S. and Malim, M.H. (2003) 'DNA Deamination Mediates Innate Immunity to Retroviral Infection', *Cell,* 113(6), pp. 803-809. doi: 10.1016/S0092-8674(03)00423-9.

Hashida, M., Kawakami, S. and Yamashita, F. (2005) 'Lipid Carrier Systems for Targeted Drug and Gene Delivery', *Chemical & Pharmaceutical Bulletin,* 53(8), pp. 871-880. doi: 10.1248/cpb.53.871.

Helseth, E., Kowalski, M., Gabuzda, D., Olshevsky, U., Haseltine, W. and Sodroski, J. (1990) 'Rapid complementation assays measuring replicative potential of human immunodeficiency virus type 1 envelope glycoprotein mutants', *Journal of Virology,* 64(5), pp. 2416-2420. doi: 10.1128/JVI.64.5.2416-2420.1990.

Herweijer, H. and Wolff, J.A. (2003) 'Progress and prospects: naked DNA gene transfer and therapy', *Gene Therapy,* 10(6), pp. 453-458. doi: 10.1038/sj.gt.3301983.

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. and Thrasher, A.J. (2008) 'Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients', *The Journal of Clinical Investigation,* 118(9), pp. 3143-3150. doi: 10.1172/JCI35798.

Huang, M.T.F. and Gorman, C.M. (1990) 'The simian virus 40 small-t intron, present in many common expression vectors, leads to aberrant splicing', *Molecular and Cellular Biology,* 10(4), pp. 1805-1810. doi: 10.1128/MCB.10.4.1805.

Huthoff, H., Bugala, K., Barciszewski, J. and Berkhout, B. (2003) 'On the importance of the primer activation signal for initiation of tRNAlys3-primed reverse transcription of the HIV-1 RNA genome', *Nucleic Acids Research,* 31(17), pp. 5186-5194. doi: 10.1093/nar/gkg714.

Hutson, T.H., Foster, E., Moon, L.D.F. and Yáñez-Muñoz, R.J. (2014) 'Lentiviral Vector-Mediated RNA Silencing in the Central Nervous System', *Human Gene Therapy,* 25(1), pp. 14-32. doi: 10.1089/hgtb.2013.016.

Ikeda, Y., Takeuchi, Y., Martin, F., Cosset, F., Mitrophanous, K. and Collins, M. (2003) 'Continuous high-titer HIV-1 vector production', *Nature Biotechnology,* 21(5), pp. 569-572. doi: 10.1038/nbt815.

Ishida, T., Obata, Y., Ohara, N., Matsushita, H., Sato, S., Uenaka, A., Saika, T., Miyamura, T., Chayama, K., Nakamura, Y., Wada, H., Yamashita, T., Morishima, T., Old, L.J. and Nakayama, E. (2008) 'Identification of the HERV-K gag antigen in prostate cancer by SEREX using autologous patient serum and its immunogenicity', *Cancer Immunity,* 8(1), pp. 15.

Isomura, H. and Stinski, M.F. (2003) 'The Human Cytomegalovirus Major Immediate-Early Enhancer Determines the Efficiency of Immediate-Early Gene Transcription and Viral Replication in Permissive Cells at Low Multiplicity of Infection', *Journal of Virology,* 77(6), pp. 3602-3614. doi: 10.1128/JVI.77.6.3602-3614.2003.

Iwakuma, T., Cui, Y. and Chang, L. (1999) 'Self-Inactivating Lentiviral Vectors with U3 and U5 Modifications', *Virology,* 261(1), pp. 120-132. doi: 10.1006/viro.1999.9850.

Jainchill, J.L., Aaronson, S.A. and Todaro, G.J. (1969) 'Murine Sarcoma and Leukemia Viruses: Assay Using Clonal Lines of Contact-Inhibited Mouse Cells', *Journal of Virology,* 4(5), pp. 549-553.

Jern, P. and Coffin, J.M. (2008) 'Effects of Retroviruses on Host Genome Function', *Annual Review of Genetics,* 42(1), pp. 709-732. doi: 10.1146/annurev.genet.42.110807.091501.

Jern, P., Sperber, G.O., Ahlsén, G. and Blomberg, J. (2005) 'Sequence Variability, Gene Structure, and Expression of Full-Length Human Endogenous Retrovirus H', *Journal of Virology,* 79(10), pp. 6325-37. doi: 10.1128/JVI.79.10.6325-6337.

Jin, L., Zeng, X., Liu, M., Deng, Y. and He, N. (2014) 'Current Progress in Gene Delivery Technology Based on Chemical Methods and Nano-carriers', *Theranostics,* 4(3), pp. 240-255. doi: 10.7150/thno.6914.

Johnson, G.E. (2011) 'Mammalian Cell HPRT Gene Mutation Assay: Test Methods'*Genetic Toxicology* New York, NY: Springer New York, pp. 55-67.

Josephson, C.D. and Abshire, T.C. (2006) 'Clinical uses of plasma and plasma fractions: plasma-derived products for hemophilias A and B, and for von Willebrand disease', *Best Practice & Research. Clinical Haematology,* 19(1), pp. 35-49. doi: 10.1016/j.beha.2005.01.031.

Kang, E.M. and Malech, H.L. (2008) 'Advances in treatment for chronic granulomatous disease', *Immunologic Research,* 43(1-3), pp. 77-84. doi: 10.1007/s12026-008-8051-z.

Kay, M.A., Glorioso, J.C. and Naldini, L. (2001) 'Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics', *Nature Medicine,* 7(1), pp. 33-40. doi: 10.1038/83324.

Kerzendorfer, C. and O'Driscoll, M. (2009) 'Human DNA damage response and repair deficiency syndromes: Linking genomic instability and cell cycle checkpoint proficiency', *DNA Repair,* 8(9), pp. 1139-1152.

Khalfallah, Y. and Genge, A. (2021) 'HERV-K inactive or potential pathogens from within?', *Journal of the Neurological Sciences,* 423, pp. 117359. doi: 10.1016/j.jns.2021.117359.

Kim, V.N., Mitrophanous, K., Kingsman, S.M. and Kingsman, A.J. (1998) 'Minimal Requirement for a Lentivirus Vector Based on Human Immunodeficiency Virus Type 1', *Journal of Virology,* 72(1), pp. 811-816. doi: 10.1128/JVI.72.1.811-816.1998.

King, A.A., Shaughnessy, D.T., Mure, K., Leszczynska, J., Ward, W.O., Umbach, D.M., Xu, Z., Ducharme, D., Taylor, J.A., DeMarini, D.M. and Klein, C.B. (2007) 'Antimutagenicity of cinnamaldehyde and vanillin in human cells: Global gene expression and possible role of DNA damage and repair', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 616(1), pp. 60-69. doi: 10.1016/j.mrfmmm.2006.11.022.

King, W., Patel, M.D., Lobel, L.I., Goff, S.P. and Nguyen-Huu, M.C. (1985) 'Insertion mutagenesis of embryonal carcinoma cells by retroviruses', *Science,* 228(4699), pp. 554-558. doi: 10.1126/science.3838595.

Kingsman, S.M., Mitrophanous, K. and Olsen, J.C. (2005) 'Potential oncogene activity of the woodchuck hepatitis post-transcriptional regulatory element (WPRE)', *Gene Therapy,* 12(1), pp. 3-4. doi: 10.1038/sj.gt.3302417.

Kjellman, C., Sjögren, H., Salford, L.G. and Widegren, B. (1999) 'HERV-F(XA34) Is a Full-Length Human Endogenous Retrovirus Expressed in Placental and Fetal Tissues', *Gene,* 239(1), pp. 99-107. doi: 10.1016/S0378-1119(99)00372-8.

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. and Baum, C. (2007) 'Retroviral vector insertion sites

associated with dominant hematopoietic clones mark "stemness" pathways', *Blood,* 109(5), pp. 1897-1907. doi: 10.1182/blood-2006-08-044156.

Kustikova, O.S., Modlich, U. and Fehse, B. (2009) 'Retroviral Insertion Site Analysis in Dominant Haematopoietic Clones'*Genetic Modification of Hematopoietic Stem Cells* Totowa, NJ: Humana Press, pp. 373-390.

Kustikova, O.S., Wahlers, A., Kühlcke, K., Stähle, B., Zander, A.R., Baum, C. and 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,* 102(12), pp. 3934-3937. doi: 10.1182/blood-2003-05-1424.

Landau, N.R., Page, K.A. and Littman, D.R. (1991) 'Pseudotyping with Human T-Cell Leukemia Virus Type I Broadens the Human Immunodeficiency Virus Host Range', *Journal of Virology,* 65(1), pp. 162-169. doi: 10.1128/JVI.65.1.162-169.1991.

Laspia, M.F., Rice, A.P. and Mathews, M.B. (1989) 'HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation', *Cell,* 59(2), pp. 283-292. doi: 10.1016/0092-8674(89)90290-0.

Lee, Y., Wang, Q., Shuryak, I., Brenner, D.J. and Turner, H.C. (2019) 'Development of a high-throughput γ-H2AX assay based on imaging flow cytometry', *Radiation Oncology,* 14(1), pp. 150. doi: 10.1186/s13014-019-1344-7.

Lesnik, E.A., Sampath, R. and Ecker, D.J. (2002) 'Rev response elements (RRE) in lentiviruses: An RNAMotif algorithm-based strategy for RRE prediction', *Medicinal research reviews,* 22(6), pp. 617-636. doi: 10.1002/med.10027.

Levine, B.L., Humeau, L.M., Boyer, J., MacGregor, R., Rebello, T., Lu, X., Binder, G.K., Slepushkin, V., Lemiale, F., Mascola, J.R., Bushman, F.D., Dropulic, B. and June, C.H. (2006) 'Gene Transfer in Humans Using a Conditionally Replicating Lentiviral Vector', *Proceedings of the National Academy of Sciences - PNAS,* 103(46), pp. 17372-17377. doi: 10.1073/pnas.0608138103.

Li, L., Olvera, J.M., Yoder, K.E., Mitchell, R.S., Butler, S.L., Lieber, M., Martin, S.L. and Bushman, F.D. (2001) 'Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection', *The EMBO Journal,* 20(12), pp. 3272-3281. doi: 10.1093/emboj/20.12.3272.

Li, M., Radvanyi, L., Yin, B., Rycaj, K., Li, J., Chivukula, R., Lin, K., Lu, Y., Shen, J., Chang, D.Z., Li, D., Johanning, G.L. and Wang-Johanning, F. (2019) 'Correction: Downregulation of Human Endogenous Retrovirus Type K (HERV-K) Viral env RNA in Pancreatic Cancer Cells Decreases

Cell Proliferation and Tumor Growth', *Clinical Cancer Research,* 25(9), pp. 2936. doi: 10.1158/1078-0432.CCR-19-0700.

Li, Z., Düllmann, J., Ostertag, W., Kühlcke, K., Eckert, H., Fehse, B., Baum, C., Schiedlmeier, B., Schmidt, M., Von Kalle, C., Meyer, J., Forster, M., Stocking, C., Wahlers, A. and Frank, O. (2002) 'Murine Leukemia Induced by Retroviral Gene Marking', *Science (American Association for the Advancement of Science),* 296(5567), pp. 497. doi: 10.1126/science.1068893.

Lo, M., Bloom, M.L., Imada, K., Berg, M., Bollenbacher, J.M., Bloom, E.T., Kelsall, B.L. and Leonard, W.J. (1999) 'Restoration of Lymphoid Populations in a Murine Model of X-Linked Severe Combined Immunodeficiency by a Gene-Therapy Approach', *Blood,* 94(9), pp. 3027-3036. doi: 10.1182/blood.V94.9.3027.

Lui, V.W., Falo Jr, L.D. and Huang, L. (2001) 'Systemic Production of Il-12 by Naked Dna Mediated Gene Transfer: Toxicity and Attenuation of Transgene Expression in Vivo', *Journal of Gene Medicine,* 3(4), pp. 384-393. doi: 10.1002/jgm.201.

Lundstrom, K. (2023) 'Viral Vectors in Gene Therapy: Where Do We Stand in 2023?', *Viruses,* 15(3), pp. 698. doi: 10.3390/v15030698.

Lysakowski, C., Dumont, L., Tramèr, M.R. and Tassonyi, E. (2003) 'A Needle-Free Jet-Injection System with Lidocaine for Peripheral Intravenous Cannula Insertion: A Randomized Controlled Trial with Cost-Effectiveness Analysis', *Anesthesia and Analgesia,* 96(1), pp. 215-219. doi: 10.1213/00000539-200301000-00044.

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. and 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 - PNAS,* 94(22), pp. 12133-12138. doi: 10.1073/pnas.94.22.12133.

Mammen, B., Ramakrishnan, T. and Sudhakar, U. (2007) 'Principles of gene therapy', *Indian Journal of Dental Research,* 18(4), pp. 196-200. doi: 10.4103/0970-9290.35832.

Mann, R., Mulligan, R.C. and Baltimore, D. (1983) 'Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus', *Cell,* 33(1), pp. 153-159. doi: 10.1016/0092-8674(83)90344-6.

Manthorpe, M., Cornefert-Jensen, F., Hartikka, J., Felgner, J., Rundell, A., Margalith, M. and Dwarki, V. (1993) 'Gene Therapy by Intramuscular Injection of Plasmid DNA: Studies on Firefly

Luciferase Gene Expression in Mice', *Human Gene Therapy,* 4(4), pp. 419-431. doi: 10.1089/hum.1993.4.4-419.

Mao, Z., Bozzella, M., Seluanov, A. and Gorbunova, V. (2008) 'DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells', *Cell Cycle,* 7(18), pp. 2902-2906. doi: 10.4161/cc.7.18.6679.

Mariotti, L.G., Pirovano, G., Savage, K.I., Ghita, M., Ottolenghi, A., Prise, K.M. and Schettino, G. (2013) 'Use of the gamma-H2AX assay to investigate DNA repair dynamics following multiple radiation exposures', *PLoS ONE,* 8(11), pp. e79541. doi: 10.1371/journal.pone.0079541.

Medina-Kauwe, L.K., Xie, J. and Hamm-Alvarez, S. (2005) 'Intracellular trafficking of nonviral vectors', *Gene Therapy,* 12(24), pp. 1734-1751. doi: 10.1038/sj.gt.3302592.

Medline Plus (2022) *What is gene therapy?* Available at: https://medlineplus.gov/genetics/understanding/therapy/genetherapy/ (Accessed: 18/04/2023).

Miller, A.D., Jolly, D.J., Friedmann, T. and Verma, I.M. (1983) 'A Transmissible Retrovirus Expressing Human Hypoxanthine Phosphoribosyltransferase (HPRT): Gene Transfer into Cells Obtained from Humans Deficient in HPRT', *Proceedings of the National Academy of Sciences - PNAS,* 80(15), pp. 4709-4713. doi: 10.1073/pnas.80.15.4709.

Miller, A.D. and Metzger, M.J. (2011) 'APOBEC3-mediated hypermutation of retroviral vectors produced from some retrovirus packaging cell lines', *Gene therapy,* 18(5), pp. 528-530. doi: 10.1038/gt.2010.177.

Miller, A.D. and Rosman, G.J. (1989) 'Improved retroviral vectors for gene transfer and expression', *BioTechniques,* 7(9), pp. 980-990.

Miller, M.D., Farnet, C.M. and Bushman, F.D. (1997) 'Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition', *Journal of Virology,* 71(7), pp. 5382-5390. doi: 10.1128/JVI.71.7.5382-5390.1997.

Min, X., Zheng, M., Yu, Y., Wu, J., Kuang, Q., Hu, Z., Ouyang, L., Lu, S. and Zhao, M. (2022) 'Ultraviolet light induces HERV expression to activate RIG-I signalling pathway in keratinocytes', *Experimental Dermatology,* Epub ahead of print(PMID: 35332586). doi: 10.1111/exd.14568.

Mitchell, R.S., Beitzel, B.F., Schroder, A.R.W., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R. and Bushman, F.D. (2004) 'Retroviral DNA Integration: ASLV, HIV, and MLV Show Distinct Target Site Preferences', *PLoS Biology,* 2(8), pp. E234. doi: 10.1371/journal.pbio.0020234.

Miyoshi, H., Blömer, U., Takahashi, M., Gage, F.H. and Verma, I.M. (1998) 'Development of a Self-Inactivating Lentivirus Vector', *Journal of Virology,* 72(10), pp. 8150-8157. doi: 10.1128/JVI.72.10.8150-8157.1998.

Modlich, U. and Baum, C. (2009) 'Preventing and exploiting the oncogenic potential of integrating gene vectors', *The Journal of Clinical Investigation,* 119(4), pp. 755-758. doi: 10.1172/JCI38831.

Modlich, U., Bohne, J., Schmidt, M., von Kalle, C., Knöss, S., Schambach, A. and Baum, C. (2006) 'Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity', *Blood,* 108(8), pp. 2545-2553. doi: 10.1182/blood-2005-08-024976.

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. and Baum, C. (2005) 'Leukemias following retroviral transfer of multidrug resistance 1 (MDR1) are driven by combinatorial insertional mutagenesis', *Blood,* 105(11), pp. 4235-4246. doi: 10.1182/blood-2004-11-4535.

Mohiuddin, I.S. and Kang, M.H. (2019) 'DNA-PK as an Emerging Therapeutic Target in Cancer', *Frontiers in Oncology,* 9, pp. 635. doi: 10.3389/fonc.2019.00635.

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. and 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', *The Journal of Clinical Investigation,* 119(4), pp. 964-975. doi: 10.1172/JCI37630.

Morange, M. (2009) 'The Central Dogma of molecular biology', *Resonance,* 14(3), pp. 236-247. doi: 10.1007/s12045-009-0024-6.

Morgan, R.A., Dudley, M.E., Wunderlich, J.R., Hughes, M.S., Yang, J.C., Sherry, R.M., Royal, R.E., Topalian, S.L., Kammula, U.S., Restifo, N.P., Zheng, Z., Nahvi, A., de Vries, C.R., Rogers-Freezer, L.J., Mavroukakis, S.A. and Rosenberg, S.A. (2006) 'Cancer Regression in Patients After Transfer of Genetically Engineered Lymphocytes', *Science,* 314(5796), pp. 126-129. doi: 10.1126/science.1129003.

Morgan, M.A., Galla, M., Grez, M., Fehse, B. and Schambach, A. (2021) 'Retroviral gene therapy in Germany with a view on previous experience and future perspectives', *Gene Therapy,* 28(9), pp. 494-512. doi: 10.1038/s41434-021-00237-x.

Moshous, D., Callebaut, I., Chasseval, R.d., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A. and Villartay, J.d. (2001) 'Artemis, a novel DNA double-strand break repair/VJ recombination protein, is mutated in human severe combined immune defeciency', *Cell,* 105(2), pp. 177.

Murnane, J.P., Fuller, L.F. and Painter, R.B. (1985) 'Establishment and characterization of a permanent pSV ori--transformed ataxia-telangiectasia cell line', *Experimental Cell Research,* 158(1), pp. 119-126. doi: 10.1016/0014-4827(85)90437-9.

Nair, V.P., Liu, H., Ciceri, G., Jungverdorben, J., Frishman, G., Tchieu, J., Cederquist, G.Y., Rothenaigner, I., Schorpp, K., Klepper, L., Walsh, R.M., Kim, T.W., Cornacchia, D., Ruepp, A., Mayer, J., Hadian, K., Frishman, D., Studer, L. and Vincendeau, M. (2021) 'Activation of HERV-K(HML-2) disrupts cortical patterning and neuronal differentiation by increasing NTRK3', *Cell Stem Cell,* 28(9), pp. 1671-1673. doi: 10.1016/j.stem.2021.05.003.

Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M. and Trono, D. (1996) 'In Vivo Gene Delivery and Stable Transduction of Nondividing Cells by a Lentiviral Vector', *Science,* 272(5259), pp. 263-267. doi: 10.1126/science.272.5259.263.

Nam, C. and Rabbitts, T.H. (2006) 'The Role of LMO2 in Development and in T Cell Leukemia After Chromosomal Translocation or Retroviral Insertion', *Molecular Therapy,* 13(1), pp. 15-25. doi: 10.1016/j.ymthe.2005.09.010.

Nandhakumar, S., Parasuraman, S., Shanmugam, M.M., Rao, K.R., Chand, P. and Bhat, B.V. (2011) 'Evaluation of DNA damage using single-cell gel electrophoresis (Comet Assay)', *Journal of Pharmacology & Pharmacotherapeutics,* 2(2), pp. 107-111. doi: 10.4103/0976-500X.81903.

Nathwani, A.C., Davidoff, A.M. and Linch, D.C. (2005) 'A review of gene therapy for haematological disorders', *British Journal of Haematology,* 128(1), pp. 3-17. doi: 10.1111/j.1365-2141.2004.05231.x.

Nathwani, A.C., Tuddenham, E.G.D., Rangarajan, S., Rosales, C., McIntosh, J., Linch, D.C., Chowdary, P., Riddell, A., Pie, A.J., Harrington, C., O'Beirne, J., Smith, K., Pasi, J., Glader, B., Rustagi, P., Ng, C.Y.C., Kay, M.A., Zhou, J., Spence, Y., Morton, C.L., Allay, J., Coleman, J., Sleep, S., Cunningham, J.M., Srivastava, D., Basner-Tschakarjan, E., Mingozzi, F., High, K.A., Gray, J.T., Reiss, U.M., Nienhuis, A.W. and Davidoff, A.M. (2011) 'Adenovirus-Associated Virus Vector–Mediated Gene Transfer in Hemophilia B', *The New England Journal of Medicine,* 365(25), pp. 2357-2365. doi: 10.1056/NEJMoa1108046.

Neumann, E., Schaefer-Ridder, M., Wang, Y. and Hofschneider, P.H. (1982) 'Gene transfer into mouse lyoma cells by electroporation in high electric fields', *The EMBO Journal,* 1(7), pp. 841-845. doi: 10.1002/j.1460-2075.1982.tb01257.x.

Nienhuis, A.W., Dunbar, C.E. and Sorrentino, B.P. (2006) 'Genotoxicity of Retroviral Integration In Hematopoietic Cells', *Molecular Therapy,* 13(6), pp. 1031-1049. doi: 10.1016/j.ymthe.2006.03.001.

Nott, A., Le Hir, H. and Moore, M.J. (2004) 'Splicing enhances translation in mammalian cells: an additional function of the exon junction complex', *Genes & Development,* 18(2), pp. 210-222. doi: 10.1101/gad.1163204.

Nowrouzi, A., Cheung, W.T., Li, T., Zhang, X., Arens, A., Paruzynski, A., Waddington, S.N., Osejindu, E., Reja, S., von Kalle, C., Wang, Y., Al-Allaf, F., Gregory, L., Themis, M., Holder, M., Dighe, N., Ruthe, A., Buckley, S.M., Bigger, B., Montini, E., Thrasher, A.J., Andrews, R., Roberts, T.P., Newbold, R.F., Coutelle, C., Schmidt, M. and Themis, M. (2013) 'The Fetal Mouse Is a Sensitive Genotoxicity Model That Exposes Lentiviral-associated Mutagenesis Resulting in Liver Oncogenesis', *Molecular Therapy,* 21(2), pp. 324-337. doi: 10.1038/mt.2012.224.

OECD (2016) *Test No. 476: In Vitro Mammalian Cell Gene Mutation Tests using the Hprt and xprt genes.* Paris: OECD (Organisation for Economic Co-operation and Development) Publishing.

Ogata, T., Okui, N., Sakuma, R., Kobayashi, N. and Kitamura, Y. (1999) 'Integrase of human endogenous retrovirus K-10 supports the replication of replication-incompetent Int- human immunodeficiency virus type 1 mutant', *Japanese Journal of Infectious Diseases,* 52(6), pp. 251-252.

O'Sullivan, B.P. and Freedman, S.D. (2009) 'Cystic fibrosis', *Lancet (London, England),* 373(9678), pp. 1891-1904. doi: 10.1016/S0140-6736(09)60327-5.

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. and 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,* 12(4), pp. 401-409. doi: 10.1038/nm1393.

Page, K.A., Landau, N.R. and Littman, D.R. (1990) 'Construction and use of a human immunodeficiency virus vector for analysis of virus infectivity', *Journal of Virology,* 64(11), pp. 5270-5276. doi: 10.1128/JVI.64.11.5270-5276.1990.

Parker, A.L., Newman, C., Briggs, S., Seymour, L. and Sheridan, P.J. (2003) 'Nonviral Gene Delivery: Techniques and Implications for Molecular Medicine', *Expert Reviews in Molecular Medicine,* 5(22), pp. 1-15. doi: 10.1017/S1462399403006562.

Patience, C., Takeuchi, Y., Cosset, F.L. and Weiss, R.A. (1998) 'Packaging of Endogenous Retroviral Sequences in Retroviral Vectors Produced by Murine and Human Packaging Cells', *Journal of Virology,* 72(4), pp. 2671-2676. doi: 10.1128/JVI.72.4.2671-2676.1998.

Perez, L.G., Davis, G.L. and Hunter, E. (1987) 'Mutants of the Rous sarcoma virus envelope glycoprotein that lack the transmembrane anchor and cytoplasmic domains: analysis of intracellular transport and assembly into virions', *Journal of Virology,* 61(10), pp. 2981-2988. doi: 10.1128/JVI.61.10.2981-2988.1987.

Pfeifer, A. and Verma, I.M. (2001) 'GENE THERAPY: Promises and Problems', *Annual Review of Genomics and Human Genetics,* 2(1), pp. 177-211. doi: 10.1146/annurev.genom.2.1.177.

Philippe, S., Sarkis, C., Barkats, M., Mammeri, H., Ladroue, C., Petit, C., Mallet, J. and Serguera, C. (2006) 'Lentiviral Vectors with a Defective Integrase Allow Efficient and Sustained Transgene Expression in vitro and in vivo', *Proceedings of the National Academy of Sciences - PNAS,* 103(47), pp. 17684-17689. doi: 10.1073/pnas.0606197103.

Phillips, G.J. (2001) 'Green Fluorescent Protein--a Bright Idea for the Study of Bacterial Protein Localization ', *FEMS Microbiology Letters,* 204(1), pp. 9-18. doi: 10.1111/j.1574-6968.2001.tb10854.x.

Poletti, V. and Mavilio, F. (2018) 'Interactions between Retroviruses and the Host Cell Genome', *Molecular Therapy—Methods & Clinical Development,* 8(C), pp. 31-41. doi: 10.1016/j.omtm.2017.10.001.

Pollard, V.W. and Malim, M.H. (1998) 'THE HIV-1 REV PROTEIN', *Annual Review of Microbiology,* 52(1), pp. 491-532. doi: 10.1146/annurev.micro.52.1.491.

Potash, A.E., Wallen, T.J., Karp, P.H., Ernst, S., Moninger, T.O., Gansemer, N.D., Stoltz, D.A., Zabner, J. and Chang, E.H. (2013) 'Adenoviral Gene Transfer Corrects the Ion Transport Defect in the Sinus Epithelia of a Porcine CF Model', *Molecular Therapy,* 21(5), pp. 947-953. doi: 10.1038/mt.2013.49.

Prasad, T.K. and Rao, N.M. (2005) 'The role of plasmid constructs containing the SV40 DNA nuclear-targeting sequence in cationic lipid-mediated DNA delivery', *Cellular & Molecular Biology Letters,* 10(2), pp. 203-215.

Pruss, D., Bushman, F.D. and Wolffe, A.P. (1994) 'Human Immunodeficiency Virus Integrase Directs Integration to Sites of Severe DNA Distortion Within the Nucleosome Core', *Proceedings of the National Academy of Sciences - PNAS,* 91(13), pp. 5913-5917. doi: 10.1073/pnas.91.13.5913.

Qasim, W., Gaspar, H.B. and Thrasher, A.J. (2009) 'Progress and prospects: gene therapy for inherited immunodeficiencies', *Gene Therapy,* 16(11), pp. 1285-1291. doi: 10.1038/gt.2009.127.

Ramamoorth, M. and Narvekar, A. (2015) 'Non Viral Vectors in Gene Therapy- An Overview', *Journal of Clinical and Diagnostic Research,* 9(1), pp. GE01-GE06. doi: 10.7860/JCDR/2015/10443.5394.

Ramdas, P., Sahu, A.K., Mishra, T., Bhardwaj, V. and Chande, A. (2020) 'From Entry to Egress: Strategic Exploitation of the Cellular Processes by HIV-1', *Frontiers in Microbiology,* 11, pp. 559792. doi: 10.3389/fmicb.2020.559792.

Ramezani, A., Hawley, T.S. and Hawley, R.G. (2008) 'Reducing the Genotoxic Potential of Retroviral Vectors'*Gene Therapy Protocols* Totowa, NJ: Humana Press, pp. 183-203.

Raper, S.E., Chirmule, N., Lee, F.S., Wivel, N.A., Bagg, A., Gao, G., Wilson, J.M. and Batshaw, M.L. (2003) 'Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer', *Molecular Genetics and Metabolism,* 80(1), pp. 148-158. doi: 10.1016/j.ymgme.2003.08.016.

Rasmussen, H., Geny, C., Deforges, L., Perron, H., Tourtelotte, W., Heltberg, A. and Clausen, J. (1995) 'Expression of endogenous retroviruses in blood mononuclear cells and brain tissue from multiple sclerosis patients', *Multiple Sclerosis,* 1(2), pp. 82-87. doi: 10.1177/135245859500100205.

Rayaprolu, V., Ramsey, J., Wang, J.C. and Mukhopadhyay, S. (2018) 'Alphavirus Purification Using Low-Speed Spin Centrifugation', *Bio-Protocol,* 8(6), pp. e2772. doi: 10.21769/BioProtoc.2772.

Redon, C.E., Dickey, J.S., Bonner, W.M. and Sedelnikova, O.A. (2009) 'γ-H2AX as A Biomarker of DNA Damage Induced by Ionizing Radiation in Human Peripheral Blood Lymphocytes and Artificial Skin', *Advances in space research,* 43(8), pp. 1171-1178. doi: 10.1016/j.asr.2008.10.011.

Rémy, S., Tesson, L., Ménoret, S., Usal, C., Scharenberg, A.M. and Anegon, I. (2009) 'Zinc-finger nucleases: a powerful tool for genetic engineering of animals', *Transgenic Research,* 19(3), pp. 363-371. doi: 10.1007/s11248-009-9323-7.

Ren, S., Li, M., Smith, J.M., DeTolla, L.J. and Furth, P.A. (2002) 'Low-volume jet injection for intradermal immunization in rabbits', *BMC Biotechnology,* 2(1), pp. 10. doi: 10.1186/1472-6750-2-10.

Richardson, J.H., Child, L.A. and Lever, A.M.L. (1993) 'Packaging of human immunodeficiency virus type 1 RNA requires cis-acting sequences outside the 5' leader region', *Journal of Virology,* 67(7), pp. 3997-4005. doi: 10.1128/JVI.67.7.3997-4005.1993.

Rieke, W.O. (1962) 'The in vivo Reutilization of Lymphocytic and Sarcoma DNA by Cells Growing in the Peritoneal Cavity', *The Journal of Cell Biology,* 13(2), pp. 205-216. doi: 10.1083/jcb.13.2.205.

Rio, D.C., Clark, S.G. and Tjian, R. (1985) 'A mammalian host-vector system that regulates expression and amplification of transfected genes by temperature induction', *Science (New York, N.Y.),* 227(4682), pp. 23-28. doi: 10.1126/science.2981116.

Roebuck, K.A. and Saifuddin, M. (1999) 'Regulation of HIV-1 Transcription', *Gene Expression,* 8(2), pp. 67-84. Obtained From: https://cognizantcommunication.com/publication/gene-expression-the-journal-of-liver-research/

Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R. and Breindl, M. (1987) 'Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites', *Journal of Virology,* 61(2), pp. 336-343. doi: 10.1128/JVI.61.2.336-343.1987.

Romani, B., Engelbrecht, S. and Glashoff, R.H. (2010) 'Functions of Tat: the versatile protein of human immunodeficiency virus type 1', *Journal of general virology,* 91(1), pp. 1-12. doi: 10.1099/vir.0.016303-0.

Romano, G. and Giordano, A. (2008) 'Role of the cyclin-dependent kinase 9-related pathway in mammalian gene expression and human diseases', *Cell Cycle,* 7(23), pp. 3664-3668. doi: 10.4161/cc.7.23.7122.

Rosenberg, S.A., Aebersold, P., Cornetta, K., Kasid, A., Morgan, R.A., Moen, R., Karson, E.M., Lotze, M.T., Yang, J.C., Topalian, S.L. and et al (1990) 'Gene transfer into humans--immunotherapy of patients with advanced melanoma, using tumor-infiltrating lymphocytes modified by retroviral gene transduction', *The New England Journal of Medicine,* 323(9), pp. 570-578. doi: 10.1056/NEJM199008303230904.

Rosonina, E., Ip, J.Y.Y., Calarco, J.A., Bakowski, M.A., Emili, A., McCracken, S., Tucker, P., Ingles, C.J. and Blencowe, B.J. (2005) 'Role for PSF in Mediating Transcriptional Activator-Dependent Stimulation of Pre-mRNA Processing In Vivo', *Molecular and Cellular Biology,* 25(15), pp. 6734-6746. doi: 10.1128/MCB.25.15.6734-6746.2005.

Rubinsky, B. (2007) 'Irreversible electroporation in medicine', *Technology in Cancer Research & Treatment,* 6(4), pp. 255-260. doi: 10.1177/153303460700600401.

Rulli, S.J., Hibbert, C.S., Mirro, J., Pederson, T., Biswal, S. and Rein, A. (2007) 'Selective and Nonselective Packaging of Cellular RNAs in Retrovirus Particles', *Journal of Virology,* 81(12), pp. 6623-6631. doi: 10.1128/JVI.02833-06.

Ruscetti, S.K. (1995) 'Erythroleukaemia induction by the Friend spleen focus-forming virus', *Baillière's clinical haematology,* 8(1), pp. 225-247. doi: 10.1016/s0950-3536(05)80239-2

Saini, S.K., Ørskov, A.D., Bjerregaard, A., Unnikrishnan, A., Holmberg-Thydén, S., Borch, A., Jensen, K.V., Anande, G., Bentzen, A.K., Marquard, A.M., Tamhane, T., Treppendahl, M.B.,

Gang, A.O., Dufva, I.H., Szallasi, Z., Ternette, N., Pedersen, A.G., Eklund, A.C., Pimanda, J., Grønbæk, K. and Hadrup, S.R. (2020) 'Human endogenous retroviruses form a reservoir of T cell targets in hematological cancers', *Nature Communications,* 11(1), pp. 5660. doi: 10.1038/s41467-020-19464-8.

Sakai, H., Siomi, H., Shida, H., Shibata, R., Kiyomasu, T. and Adachi, A. (1990) 'Functional comparison of transactivation by human retrovirus rev and rex genes', *Journal of Virology,* 64(12), pp. 5833-5839. doi: 10.1128/JVI.64.12.5833-5839.1990.

Sakuma, T., Barry, M.A. and Ikeda, Y. (2012) 'Lentiviral vectors: Basic to translational', *Biochemical Journal,* 443(3), pp. 603-618. doi: 10.1042/BJ20120146.

Sari, U. and Tiryaki, I. (2022) 'Genetic Transformation Techniques', in Yilmaz, T. (ed.) *Engineering and Architecture Sciences: Theory, Current Researches and New Trends.* 1st edn. Cetinje: IVPE, pp. 156.

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. and von Kalle, C. (2002) 'Polyclonal long-term repopulating stem cell clones in a primate model', *Blood,* 100(8), pp. 2737-2743. doi: 10.1182/blood-2002-02-0407.

Schneider, R., Campbell, M., Nasioulas, G., Felber, B.K. and Pavlakis, G.N. (1997) 'Inactivation of the human immunodeficiency virus type 1 inhibitory elements allows Rev-independent expression of Gag and Gag/protease and particle formation', *Journal of Virology,* 71(7), pp. 4892-4903. doi: 10.1128/JVI.71.7.4892-4903.1997.

Scholl, H.P.N. and Sahel, J.A. (2014) 'Gene therapy arrives at the macula', *The Lancet,* 383(9923), pp. 1105-1107. doi: 10.1016/S0140-6736(14)60033-7.

Schröder, A.R.W., Shinn, P., Chen, H., Berry, C., Ecker, J.R. and Bushman, F. (2002) 'HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots', *Cell,* 110(4), pp. 521-529. doi: 10.1016/S0092-8674(02)00864-4.

Scollay, R. (2001) 'Gene therapy: a brief overview of the past, present, and future', *Annals of the New York Academy of Sciences,* 953(1), pp. 26-30. doi: 10.1111/j.1749-6632.2001.tb11357.x.

Selkirk, S.M. (2004) 'Gene therapy in clinical medicine', *Postgraduate Medical Journal,* 80(948), pp. 560-570. doi: 10.1136/pgmj.2003.017764.

Shaw, G., Morse, S., Ararat, M. and Graham, F.L. (2002) 'Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells', *Faseb Journal: Official Publication of The Federation of American Societies for Experimental Biology,* 16(8), pp. 869-871. doi: 10.1096/fj.01-0995fje.

Shcherbakova, D.M. and Verkhusha, V.V. (2013) 'Near-infrared fluorescent proteins for multicolor in vivo imaging', *Nature Methods,* 10(8), pp. 751-754. doi: 10.1038/nmeth.2521.

Skalka, A.M. and Katz, R.A. (2005) 'Retroviral DNA integration and the DNA damage response', *Cell Death and Differentiation,* 12(S1), pp. 971-978. doi: 10.1038/sj.cdd.4401573.

Smart, D.J., Halicka, H.D., Schmuck, G., Traganos, F., Darzynkiewicz, Z. and Williams, G.M. (2008) 'Assessment of DNA double-strand breaks and γH2AX induced by the topoisomerase II poisons etoposide and mitoxantrone', *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis,* 641(1-2), pp. 43-47. doi: 10.1016/j.mrfmmm.2008.03.005.

Smith, J.K., Cywinski, A. and Taylor, J.M. (1984) 'Specificity of initiation of plus-strand DNA by Rous sarcoma virus', *Journal of Virology,* 52(2), pp. 314-319. doi: 10.1128/JVI.52.2.314-319.1984.

Solbak, S.M.Ø, Reksten, T.R., Hahn, F., Wray, V., Henklein, P., Henklein, P., Halskau, Ø, Schubert, U. and Fossen, T. (2013) 'HIV-1 p6 — a structured to flexible multifunctional membrane-interacting protein', *Biochimica et Biophysica Acta (BBA) - Biomembranes,* 1828(2), pp. 816-823. doi: 10.1016/j.bbamem.2012.11.010.

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. and Grez, M. (2010) 'Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease', *Nature Medicine,* 16(2), pp. 198-204. doi: 10.1038/nm.2088.

Stein, U., Walther, W., Stege, A., Kaszubiak, A., Fichtner, I. and Lage, H. (2008) 'Complete In Vivo Reversal of the Multidrug Resistance Phenotype by Jet-injection of Anti-MDR1 Short Hairpin RNA-encoding Plasmid DNA', *Molecular Therapy,* 16(1), pp. 178-186. doi: 10.1038/sj.mt.6300304.

Steinhuber, S., Brack, M., Hunsmann, G., Schwelberger, H., Dierich, M.P. and Vogetseder, W. (1995) 'Distribution of human endogenous retrovirus HERV-K genomes in humans and different primates', *Human Genetics,* 96(2), pp. 188-192. doi: 10.1007/BF00207377.

Stocking, C., Bergholz, U., Friel, J., Klingler, K., Wagener, T., Starke, C., Kitamura, T., Miyajima, A. and Ostertag, W. (1993) 'Distinct Classes of Factor-Independent Mutants can be Isolated after Retroviral Mutagenesis of a Human Myeloid Stem Cell Line', *Growth Factors,* 8(3), pp. 197-209. doi: 10.3109/08977199309011023.

Stoye, J.P. (2012) 'Studies of endogenous retroviruses reveal a continuing evolutionary saga', *Nature Reviews Microbiology,* 10(6), pp. 395-406. doi: 10.1038/nrmicro2783.

Tan, Y., Whitmore, M., Li, S., Frederik, P. and Huang, L. (2002) 'LPD nanoparticles--novel nonviral vector for efficient gene delivery', *Methods in Molecular Medicine,* 69, pp. 73-81. doi: 10.1385/1-59259-141-8:073

Tarakanova, V.L., Leung-Pineda, V., Hwang, S., Yang, C., Matatall, K., Basson, M., Sun, R., Piwnica-Worms, H., Sleckman, B.P. and Virgin, 4., Herbert W (2007) 'γ-Herpesvirus Kinase Actively Initiates a DNA Damage Response by Inducing Phosphorylation of H2AX to Foster Viral Replication', *Cell Host & Microbe,* 1(4), pp. 275-286. doi: 10.1016/j.chom.2007.05.008.

Tarantal, A.F., McDonald, R.J., Jimenez, D.F., Lee, C.C.I., O'Shea, C.E., Leapley, A.C., Won, R.H., Plopper, C.G., Lutzko, C. and 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,* 12(1), pp. 87-98. doi: 10.1016/j.ymthe.2005.01.019.

Taya, M. and Kino-oka, M. (2011) '2.27 - Bioreactors for Animal Cell Cultures', in Moo-Young, M. (ed.) *Comprehensive Biotechnology (Second Edition)* Burlington: Academic Press, pp. 373-382.

Terwilliger, E., Burghoff, R., Sia, R., Sodroski, J., Haseltine, W. and Rosen, C. (1988) 'The art gene product of human immunodeficiency virus is required for replication', *Journal of Virology,* 62(2), pp. 655-658. doi: 10.1128/JVI.62.2.655-658.1988.

Themis, M., May, D., Coutelle, C. and 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,* 10(19), pp. 1703-1711. doi: 10.1038/sj.gt.3302059.

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. and Coutelle, C. (2005) 'Oncogenesis Following Delivery of a Nonprimate Lentiviral Gene Therapy Vector to Fetal and Neonatal Mice', *Molecular Therapy,* 12(4), pp. 763-771. doi: 10.1016/j.ymthe.2005.07.358.

Thomas, C.E., Ehrhardt, A. and Kay, M.A. (2003) 'Progress and problems with the use of viral vectors for gene therapy', *Nature Reviews Genetics,* 4(5), pp. 346-358. doi: 10.1038/nrg1066.

Thompson, A.J.V. and Patel, K., MD (2009) 'Antisense Inhibitors, Ribozymes, and siRNAs', *Clinics in Liver Disease,* 13(3), pp. 375-390. doi: 10.1016/j.cld.2009.05.003.

Thrasher, A.J., Baum, C., Candotti, F., Blyth, K., Cavazzana-Calvo, M., Sorrentino, B., Modlich, U., Neil, J., Schambach, A., Cameron, E., Otsu, M., Fischer, A., Gaspar, H.B., Scobie, L. and

Abina, S.H. (2006) 'Gene therapy X-SCID transgene leukaemogenicity', *Nature,* 443(7109), pp. E5-E6. doi: 10.1038/nature05219.

Tian, F., Xu, G., Zhou, S., Chen, S. and He, D. (2023) 'Principles and Applications of Green Fluorescent Protein-based Biosensors: a Mini-Review', *Analyst,* (published online ahead of print 31 May). Available at doi: 10.1039/d3an00320e.

Tipper, C.H., Cingöz, O. and Coffin, J.M. (2012) 'Mus spicilegus Endogenous Retrovirus HEMV Uses Murine Sodium-Dependent Myo-Inositol Transporter 1 as a Receptor', *Journal of Virology,* 86(11), pp. 6341-6344. doi: 10.1128/JVI.00083-12.

Toyoda, T and Ogawa, K. (2022) 'Early detection of urinary bladder carcinogens in rats by immunohistochemistry for γ-H2AX: a review from analyses of 100 chemicals', *Journal of Toxicologic Pathology,* 35(4), pp. 283-298. doi: 10.1293/tox.2022-0061.

Toyoda, T., Sone, M., Matsushita, K., Akane, H., Akagi, J., Morikawa, T., Mizuta, Y., Cho, Y. and Ogawa, K. (2023) 'Early detection of hepatocarcinogens in rats by immunohistochemistry of γ-H2AX', *Journal of Toxicological Sciences,* 48(6), pp. 323-332. doi: 10.2131/jts.48.323.

Toyooka, T., Ishihama, M. and Ibuki, Y. (2011) 'Phosphorylation of Histone H2AX Is a Powerful Tool for Detecting Chemical Photogenotoxicity', *Journal of Investigative Dermatology,* 131(6), pp. 1313-1321. doi: 10.1038/jid.2011.28.

Tsien, R.Y. (1998) 'The Green Fluorescent Protein', *Annual Review of Biochemistry,* 67(1), pp. 509-544. doi: 10.1146/annurev.biochem.67.1.509.

Tsuji, T., Itoh, K., Baum, C., Ohnishi, N., Tomiwa, K., Hirano, D., Nishimura-Morita, Y., Ostertag, W. and Fujita, J. (2000) 'Retroviral vector-mediated gene expression in human CD34+CD38− cells expanded in vitro: cis elements of FMEV are superior to those of Mo-MuLV', *Human Gene Therapy,* 11(2), pp. 271-284. doi: 10.1089/10430340050016012.

Turner, G., Barbulescu, M., Su, M., Jensen-Seaman, M.I., Kidd, K.K. and Lenz, J. (2001) 'Insertional polymorphisms of full-length endogenous retroviruses in humans', *Current Biology,* 11(19), pp. 1531-1535. doi: 10.1016/S0960-9822(01)00455-9.

Ueno, N.T., Bartholomeusz, C., Xia, W., Anklesaria, P., Bruckheimer, E.M., Mebel, E., Paul, R., Li, S., Yo, G.H., Huang, L. and Hung, M. (2002) 'Systemic Gene Therapy in Human Xenograft Tumor Models by Liposomal Delivery of the E1A Gene', *Cancer Research,* 62(22), pp. 6712-6716.

Uren, A.G., Kool, J., Berns, A. and van Lohuizen, M. (2005) 'Retroviral insertional mutagenesis: past, present and future', *Oncogene,* 24(52), pp. 7656-72. doi: 10.1038/sj.onc.1209043.

van der Kuyl, Antoinette C (2012) 'HIV infection and HERV expression: a review', *Retrovirology,* 9(1), pp. 6. doi: 10.1186/1742-4690-9-6.

Varmus, H.E. (1982) 'Form and function of retroviral proviruses', *Science,* 216(4548), pp. 812-820. doi: 10.1126/science.6177038.

Verma, I.M., Meuth, N.L., Bromfeld, E., Manly, K.F. and Baltimore, D. (1971) 'Covalently Linked RNA-DNA Molecule as Initial Product of RNA Tumour Virus DNA Polymerase', *Nature New biology (London),* 233(39), pp. 131-134. doi: 10.1038/newbio233131a0.

Verma, I.M. and Weitzman, M.D. (2005) 'Gene therapy: twenty-first century medicine', *Annual Review of Biochemistry,* 74(1), pp. 711-738. doi: 10.1146/annurev.biochem.74.050304.091637.

Vigna, E. and Naldini, L. (2000) 'Lentiviral vectors: excellent tools for experimental gene transfer and promising candidates for gene therapy', *Journal of Gene Medicine,* 2(5), pp. 308-316. doi: 10.1002/1521-2254(200009/10)2:5<308::AID-JGM131>3.0.CO;2-3.

Vignard, J., Mirey, G. and Salles, B. (2013) 'Ionizing-radiation induced DNA double-strand breaks: A direct and indirect lighting up', *Radiotherapy and Oncology,* 108(3), pp. 362-369. doi: 10.1016/j.radonc.2013.06.013.

Vijaya, S., Steffen, D.L. and Robinson, H.L. (1986) 'Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin', *Journal of Virology,* 60(2), pp. 683-692. doi: 10.1128/JVI.60.2.683-692.1986.

Vincze, T., Posfai, J. and Roberts, R.J. (2003) 'NEBcutter: a program to cleave DNA with restriction enzymes', *Nucleic Acids Research,* 31(13), pp. 3688-3691. doi: 10.1093/nar/gkg526.

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., Cook, H.T., Ali, R.R., Thrasher, A., Tuddenham, E.G.D., Themis, M. and Coutelle, C. (2004) 'Permanent phenotypic correction of hemophilia B in immunocompetent mice by prenatal gene therapy', *Blood,* 104(9), pp. 2714-2721. doi: 10.1182/blood-2004-02-0627.

Wagner, J.A., Nepomuceno, I.B., Messner, A.H., Moran, M.L., Batson, E.P., Dimiceli, S., Brown, B.W., Desch, J.K., Norbash, A.M., Conrad, C.K., Guggino, W.B., Flotte, T.R., Wine, J.J., Carter, B.J., Reynolds, T.C., Moss, R.B. and Gardner, P. (2002) 'A Phase II, Double-Blind, Randomized, Placebo-Controlled Clinical Trial of tgAAVCF Using Maxillary Sinus Delivery in Patients with Cystic Fibrosis with Antrostomies', *Human gene therapy,* 13(11), pp. 1349-1359. doi: 10.1089/104303402760128577.

Walsh, C.E. (1999) 'Fetal Gene Therapy', *Gene Therapy,* 6(7), pp. 1200-1201.

Walther, W., Siegel, R., Schlag, P.M., Kobelt, D., Knösel, T., Dietel, M., Bembenek, A., Aumann, J., Schleef, M., Baier, R. and Stein, U. (2008) 'Novel Jet-Injection Technology for Nonviral Intratumoral Gene Transfer in Patients with Melanoma and Breast Cancer', *Clinical Cancer Research,* 14(22), pp. 7545-7553. doi: 10.1158/1078-0432.CCR-08-0412.

Wang, L., Takabe, K., Bidlingmaier, S.M., III, C.R. and Verma, I.M. (1999) 'Sustained Correction of Bleeding Disorder in Hemophilia B Mice by Gene Therapy', *Proceedings of the National Academy of Sciences - PNAS,* 96(7), pp. 3906-3910. doi: 10.1073/pnas.96.7.3906.

Wang, W., Li, W., Ma, N. and Steinhoff, G. (2013) 'Non-Viral Gene Delivery Methods', *Current Pharmaceutical Biotechnology,* 14(1), pp. 46-60. doi: 10.2174/1389201011314010008.

Wanisch, K. and Yáñez-Muñoz, R.J. (2009) 'Integration-deficient Lentiviral Vectors: A Slow Coming of Age', *Molecular Therapy,* 17(8), pp. 1316-1332. doi: 10.1038/mt.2009.122.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Swanstrom, R., Burch, C.L. and Weeks, K.M. (2009) 'Architecture and secondary structure of an entire HIV-1 RNA genome', *Nature,* 460(7256), pp. 711-716. doi: 10.1038/nature08237.

Weiss, R.A. (2006) 'The discovery of endogenous retroviruses', *Retrovirology,* 3(1), pp. 67. doi: 10.1186/1742-4690-3-67.

Widegren, B., Kjellman, C., Aminoff, S., Sahlford, L.G. and Sjogren, H. (1996) 'The Structure and Phylogeny of A New Family of Human Endogenous Retroviruses', *Journal of General Virology,* 77(8), pp. 1631-1641. doi: 10.1099/0022-1317-77-8-1631.

Wildschutte, J.H., Williams, Z.H., Montesion, M., Subramanian, R.P., Kidd, J.M. and Coffin, J.M. (2016) 'Discovery of unfixed endogenous retrovirus insertions in diverse human populations', *Proceedings of the National Academy of Sciences - PNAS,* 113(16), pp. E2326-E2334. doi: 10.1073/pnas.1602336113.

Wolf, D.P., Thomson, J.A., Zelinski-Wooten, M.B. and Stouffer, R.L. (1990) 'In vitro fertilization-embryo transfer in nonhuman primates: The technique and its applications', *Molecular Reproduction and Development,* 27(3), pp. 261-280. doi: 10.1002/mrd.1080270313.

Wu, X., Li, Y., Crise, B. and Burgess, S.M. (2003) 'Transcription Start Regions in the Human Genome Are Favored Targets for MLV Integration', *Science,* 300(5626), pp. 1749-1751. doi: 10.1126/science.1083413.

Wu, Y. and Marsh, J.W. (2001) 'Selective transcription and modulation of resting T cell activity by preintegrated HIV DNA', *Science (New York, N.Y.),* 293(5534), pp. 1503-1506. doi: 10.1126/science.1061548.

Yáñez-Muñoz, R.J., Ali, R.R., Balaggan, K.S., MacNeil, A., Howe, S.J., Schmidt, M., Smith, A.J., Buch, P., MacLaren, R.E., Anderson, P.N., Barker, S.E., Duran, Y., Bartholomae, C., von Kalle, C., Heckenlively, J.R., Kinnon, C. and Thrasher, A.J. (2006) 'Effective gene therapy with nonintegrating lentiviral vectors', *Nature Medicine,* 12(3), pp. 348-353. doi: 10.1038/nm1365.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T.L. (2012) 'Primer-BLAST: A Tool to Design Target-Specific Primers for Polymerase Chain Reaction', *BMC Bioinformatics,* 13(1), pp. 134. doi: 10.1186/1471-2105-13-134.

Yeung, G., Choi, L.M., Chao, L.C., Park, N.J., Liu, D., Jamil, A. and Martinson, H.G. (1998) 'Poly(A)-Driven and Poly(A)-Assisted Termination: Two Different Modes of Poly(A)-Dependent Transcription Termination', *Molecular and Cellular Biology,* 18(1), pp. 276-289. doi: 10.1128/MCB.18.1.276.

Yuan, J., Xu, W.W., Jiang, S., Yu, H. and Poon, H.F. (2018) 'The Scattered Twelve Tribes of HEK293', *Biomedical and Pharmacology Journal,* 11(2), pp. 621-623.

Zanta-Boussif, M.A., Charrier, S., Brice-Ouzet, A., Martin, S., Opolon, P., Thrasher, A.J., Hope, T.J. and Galy, A. (2009) 'Validation of a mutated PRE sequence allowing high and sustained transgene expression while abrogating WHV-X protein synthesis: application to the gene therapy of WAS', *Gene Therapy,* 16(5), pp. 605-619. doi: 10.1038/gt.2009.3.

Zeilfelder, U., Frank, O., Sparacio, S., Schön, U., Bosch, V., Seifarth, W. and Leib-Mösch, C. (2007) 'The potential of retroviral vectors to cotransfer human endogenous retroviruses (HERVs) from human packaging cell lines', *Gene,* 390(1), pp. 175-179. doi: 10.1016/j.gene.2006.08.019.

Zennou, V., Serguera, C., Sarkis, C., Colin, P., Perret, E., Mallet, J. and Charneau, P. (2001) 'The HIV-1 DNA flap stimulates HIV vector-mediated cell transduction in the brain', *Nature Biotechnology,* 19(5), pp. 446-450. doi: 10.1038/88115.

Zhang, H., Pomerantz, R.J., Dornadula, G. and Sun, Y. (2000) 'Human Immunodeficiency Virus Type 1 Vif Protein Is an Integral Component of an mRNP Complex of Viral RNA and Could Be Involved in the Viral RNA Folding and Packaging Process', *Journal of Virology,* 74(18), pp. 8252-8261. doi: 10.1128/JVI.74.18.8252-8261.2000.

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L. and Friedman, J.M. (1994) 'Positional cloning of the mouse obese gene and its human homologue', *Nature,* 372(6505), pp. 425-432. doi: 10.1038/372425a0.

Zhao, Y., Keating, K., Dolman, C. and Thorpe, R. (2008) 'Characterization of complete particles (VSV-G/SIN-GFP) and empty particles (VSV-G/EMPTY) in human immunodeficiency virus type 1-based lentiviral products for gene therapy: potential applications for improvement of product quality and safety', *Human Gene Therapy,* 19(5), pp. 475-486. doi: 10.1089/hum.2007.119.

Zuba-Surma, E.K., Kucia, M., Abdel-Latif, A., Lillard, J., James W and Ratajczak, M.Z. (2007) 'The ImageStream System: a key step to a new era in imaging', *Folia Histochemica et Cytobiologica,* 45(4), pp. 279-290. doi: 10.5603/4500.

Zufferey, R., Donello, J.E., Trono, D. and Hope, T.J. (1999) 'Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element Enhances Expression of Transgenes Delivered by Retroviral Vectors', *Journal of Virology,* 73(4), pp. 2886-2892. doi: 10.1128/JVI.73.4.2886-2892.1999.

Zufferey, R., Dull, T., Mandel, R.J., Bukovsky, A., Quiroz, D., Naldini, L. and Trono, D. (1998) 'Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery', *Journal of Virology,* 72(12), pp. 9873-9880. doi: 10.1128/JVI.72.12.9873-9880.1998.

Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L. and Trono, D. (1997) 'Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo', *Nature Biotechnology,* 15(9), pp. 871-875. doi: 10.1038/nbt0997-871.

Appendix

Appendix 1 – Previous Immunocytochemistry work carried out

In prior work carried out in our laboratory by Safia Reja, the γH2AX DSB profiles of MRC5-SV1 and the XP14BRneo17 were compared. Both cell lines are immortalised lung fibroblasts, the difference being that the MRC cell line has normal DNA repair while the XP14 cell line is defective in its NHEJ mechanism due to a lack of DNA PKCs. Immunocytochemistry was used in the form of a γH2AX assay to determine the DNA damage profiles.

The MRC5 cell line was exposed to radiation, which is also known to cause DSBs (Vignard, Mirey and Salles, 2013). Irradiation of MRC5 was compared to transduction of these cells with murine leukaemia virus (MLV) packaging a retrovirus genome and an "empty" MLV gene therapy vector, which lacked vector backbone. The DNA damage profiles were then observed over various time points.

Figure AP1 γH2AX DNA damage profile of MRC5-SV1 from previous data. γH2AX staining was done on the NHEJ competent MRC5 cells to observe the number of DSB seen utilising immunocytochemistry. Foci of γH2AX staining indicate a location with at least one DSB. Following irradiation, a high number of DSBs are seen which are repaired over time. When infected with the MLV, a much larger number of DSBs are seen as time progresses following repair. Infecting the cells with an

empty vector lacking the viral genome, there is a much higher amount of DSBs initially followed by repair. Data provided by Dr. Michael Themis, Brunel University London.
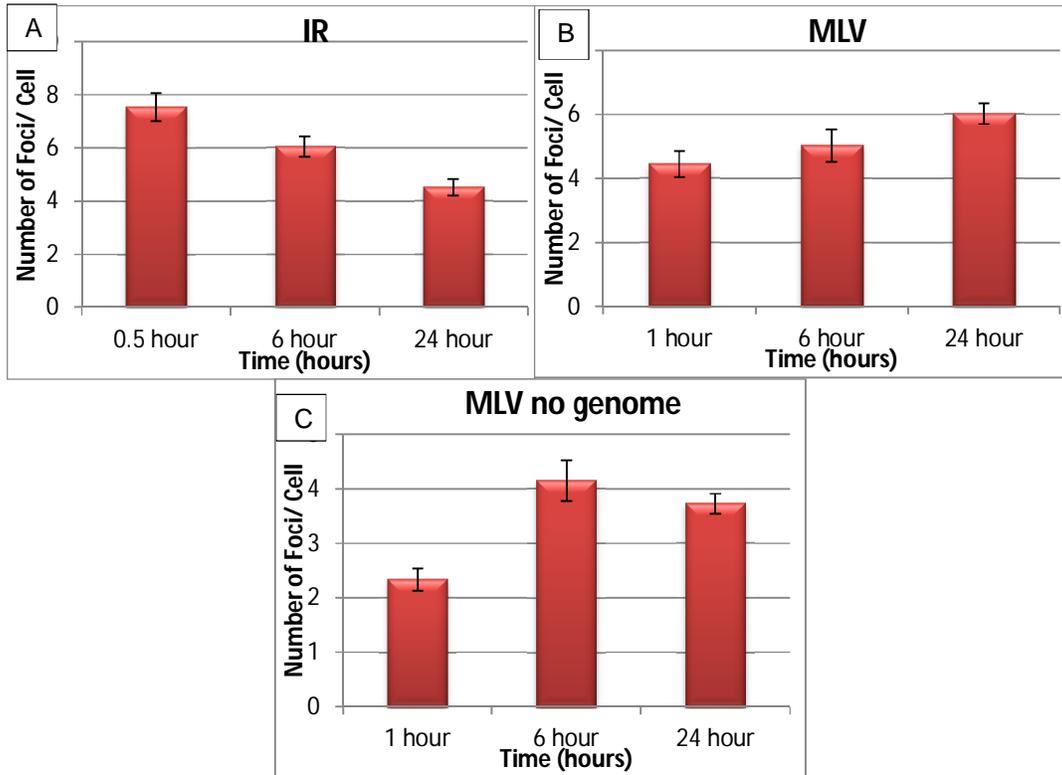
Figure AP2 γH2AX DNA damage profile of XP14BRneo17 from previous data. γH2AX staining was done on NHEJ compromised XP14 cells to observe the number of DSBs seen utilising immunocytochemistry. Foci of γH2AX staining indicate a location with at least one DSB. Following irradiation, a high number of DSB are seen which are repaired over time but at a much slower rate compared to the DNA repair normal MRC cells (**figure AP1**). When infected with the MLV no repair is seen over time. Infecting the cells with an empty vector lacking the viral genome indicated repair by the 24 hour mark, however it was still overlapping with the damage seen in the 6 hour mark if the margins of error are observed. Data provided by Dr. Michael Themis, Brunel University London, 2016.

Appendix 2 – Previous ImageStream work carried out

Prior work was conducted over more time points following MLV and HIV-F9 infection. ImageStream, an advanced flow cytometer with an imaging device which can observe a high throughput of individual cells in high resolution, was used to process the cells and count the DSBs.
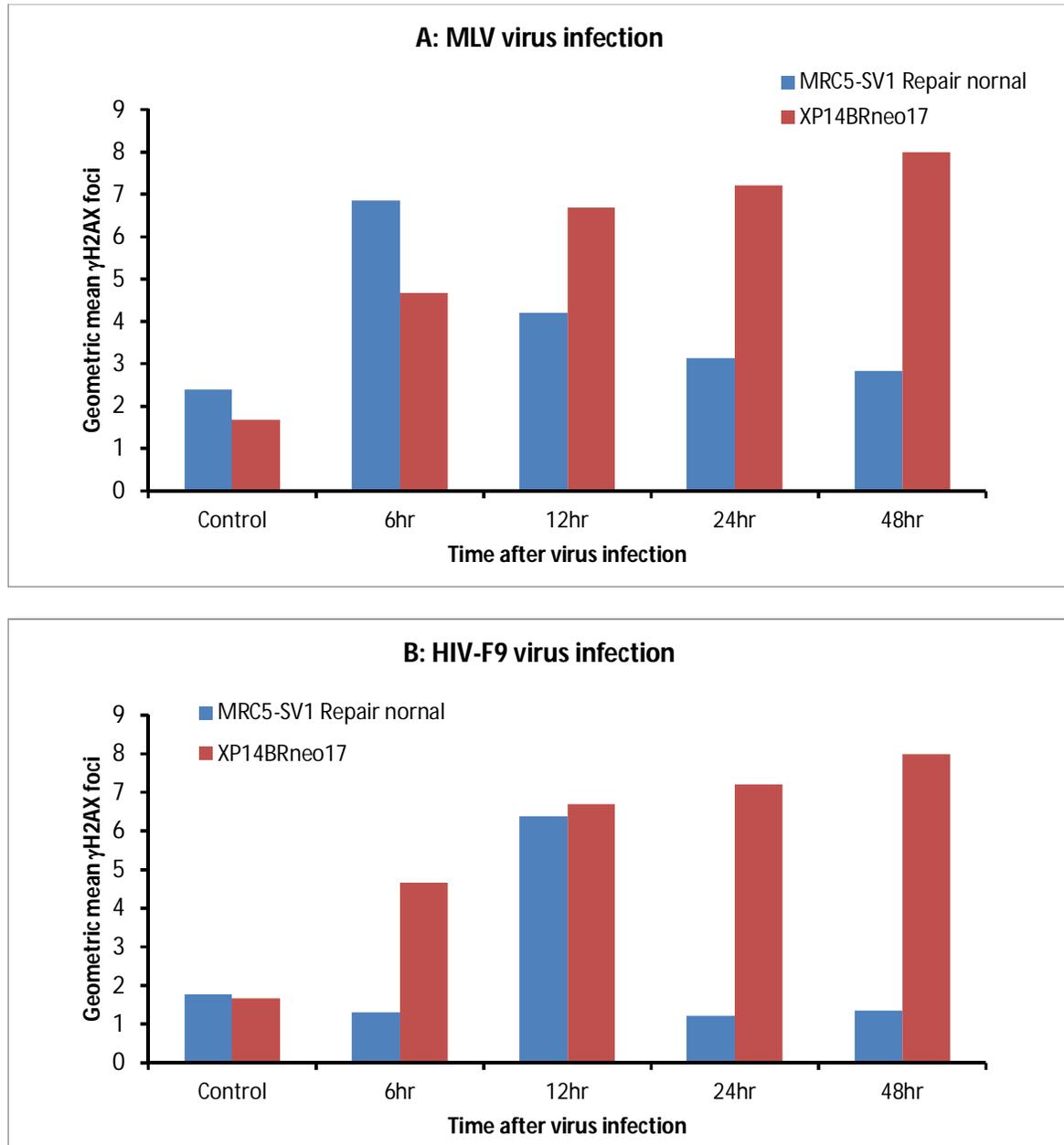
Figure AP3 γH2AX DNA damage profile of MRC5-SV1 and XP14BRneo17 following MLV and HIVF9 infection from previous data. γH2AX staining was done on the cells to observe the DNA damage profile and analysed with ImageStream analysis. The geometric mean of γH2AX staining indicate individual DSBs in a collective amount of cells. The MLV infection showed an increase in γH2AX

foci in the MRC5 cell line, peaking at six hours followed by a decline as the cells repaired the damage, however the repair was unable to restore cells back to the control levels by the 48 hour mark. The XP14 cell line showed an increase without repair up to the 48 hour mark. The HIV infection showed a sharp increase at 12 hours followed by repair showing a return to control level DNA breaks. The XP14 cell line infected with the HIV showed a similar pattern as seen with the MLV infection with a constant increase without repair.

Appendix 3 - Immunocytochemistry Raw Data

AM7 Producer Line – Human Produced MLV transduction 1

MRC5-SV1

XP14BRneo17

Time point	0	6	12	24	48	Time point	0	6	12	24	48
Total Foci	144	693	513	301	283	Total Foci	420	975	732	752	748
Cumulative Mean	1.44	6.93	5.13	3.01	2.83	Cumulative Mean	4.2	9.75	7.32	7.52	7.48
Std Error	0.20	0.70	0.43	0.40	0.57	Std Error	0.48	0.70	0.56	0.56	0.47

AM7 Producer Line – Human Produced MLV transduction 2

MRC5-SV1

XP14

Time point	0	6	12	24	48	Time point	0	6	12	24	48
Total Foci	401	1093	906	783	598	Total Foci	557	1122	1097	1057	947
Cumulative mean	4.01	10.93	9.06	7.83	5.98	Cumulative mean	5.57	11.22	10.97	10.57	9.47
Std Error	0.36	0.46	0.34	0.47	0.42	Std Error	0.38	0.46	0.36	0.36	0.38

PA317 Producer Line– Mouse Produced MLV with No Genome – Transduction 1

MRC5-SV1

XP14BRneo17

Time point	0	6	12	24	48	Time point	0	6	12	24	48
Total Foci	144	693	513	301	283	Total Foci	420	975	732	752	748
Cumulative mean	1.44	6.93	5.13	3.01	2.83	Cumulative mean	4.2	9.75	7.32	7.52	7.48
Std Error	0.20	0.70	0.43	0.40	0.57	Std Error	0.48	0.78	0.56	0.56	0.47

PA317 Producer Line– Mouse Produced MLV with No Genome – Transduction 2

MRC5-SV1   XP14

Time point	0	6	12	24	48	Time point	0	6	12	24	48
Total Foci	367	1072	869	646	618	Total Foci	461	1118	1034	1061	1041
Cumulative mean	3.67	10.72	8.69	6.46	6.18	Cumulative mean	4.61	11.18	10.34	10.61	10.41
Std Error	0.19	0.34	0.29	0.31	0.34	Std Error	0.23	0.47	0.45	0.37	0.36

PA317 Producer Line– Mouse Produced MLV with No Genome – Transduction 3

MRC5-SV1   XP14BRneo17

Time point	0	6	12	24	48	Time point	0	6	12	24	48
Total Foci	245	868	723	591	514	Total Foci	278	955	943	907	950
Cumulative mean	2.45	8.68	7.23	5.91	5.14	Cumulative mean	2.78	9.55	9.43	9.07	9.5
Std Error	0.22	0.54	0.26	0.34	0.22	Std Error	0.25	0.37	0.37	0.37	0.45

HIV LV Complete Genome – Transduction 1

MRC5-SV1   XP14BRneo17

Time point	0	6	12	24	48	Time point	0	6	12	24	48
Total Foci	245	1011	769	625	548	Total Foci	278	1094	958	939	931
Cumulative mean	2.45	10.11	7.69	6.25	5.48	Cumulative mean	2.78	10.94	9.58	9.39	9.31
Std Error	0.22	0.57	0.37	0.31	0.24	Std Error	0.25	0.44	0.41	0.46	0.35

HIV LV Complete Genome – Transduction 2

MRC5-SV1   XP14BRneo17

Time point	0	6	12	24	48	Time point	0	6	12	24	48
Total Foci	319	1110	728	672	511	Total Foci	232	949	890	874	753
Cumulative mean	3.19	11.1	7.28	6.72	5.11	Cumulative mean	2.32	9.49	8.9	8.83	7.68

Std Error	0.18	0.48	0.23	0.24	0.25	Std Error	0.22	0.44	0.36	0.26	0.33

HIV LV Empty Genome – Transduction 1

MRC5-
SV1   XP14BRneo17

Time point	0	6	12	24	48	Time point	0	6	12	24	48
Total Foci	319	891	754	661	558	Total Foci	232	1128	845	862	871
Cumulative mean	3.19	8.91	7.54	6.61	5.58	Cumulative mean	2.32	11.28	8.45	8.62	8.71
Std Error	0.18	0.33	0.33	0.26	0.34	Std Error	0.22	0.95	0.28	0.25	0.37

HIV LV Empty Genome – Transduction 2

MRC5-SV1   XP14BRneo17

Time point	0	6	12	24	48	Time point	0	6	12	24	48
Total Foci	319	897	763	650	540	Total Foci	232	946	828	835	848
Cumulative mean	3.19	8.97	7.63	6.5	5.4	Cumulative mean	2.32	9.46	8.28	8.35	8.48
Std Error	0.18	0.52	0.24	0.35	0.35	Std Error	0.22	0.50	0.40	0.40	0.40

Appendix 4 – HERV K113 Sequence and Primer Pairs

AP4.1 HERV Sequence Primers

```
Total Sequence: 9472bp
LTR1: 968bp || LTR2:969bp
Region between LTRs: 7535bp
```

```
Yellow highlights indicate LTR sequences in the total HERV region.
```

```
Other coloured highlights indicate primer sequence bind sites and overlaps
between primer bind sites and LTR regions as well as inter primer bind sites.
```

<mark style="background:yellow">LTR Regions</mark>

Primer Pair 1
```
CAACCCACCCCTACATCTGG
CGACAAAACCACCATCGTCA
```

<mark style="background:#00ff00">Forward and Reverse Primer</mark>
<mark style="background:cyan">LTR and Forward Primer Overlap</mark>
<mark style="background:red">Reverse Primer 1+4 Overlap</mark>

```
PCR Product: 7534bp
```

Primer Pair 2
```
TCTCTAGGGTGAAGGTACGC
TGCTTTTCCCCACATTTCCC
```

<mark style="background:#999999">Forward and Reverse Primer</mark>
<mark style="background:teal">LTR + Reverse Primer 2+3 Overlap</mark>
<mark style="background:purple">Forward Primer 2 and 4 overlap</mark>
<mark style="background:#1a1a99">Forward Primer 2 and 3 overlap</mark>

```
PCR Product: 7526bp
```

Primer Pair 3
```
CTCGAGCGTGGTCATTGAG
CTCTCTTGCTTTTCCCCACATT
```

<mark style="background:green">Forward and Reverse Primer</mark>
<mark style="background:magenta">LTR and Reverse Primer Overlap</mark>
<mark style="background:teal">LTR + Reverse Primer 2+3 Overlap</mark>
<mark style="background:#1a1a99">Forward Primer 2 and 3 overlap</mark>

PCR Product: 7513bp

Primer Pair 4

GGCTTTTCTCTAGGGTGAAGG

GACAAAACCACCATCGTCATCA

Forward and Reverse Primer
Reverse Primer 1+4 Overlap
Forward Primer 2 and 4 overlap

PCR Product: 7500bp

Primer Pair 5

ACAAGTCGACGAGAGATCCC

CGACAAAACCACCATCGTCA

Forward and Reverse Primer: White letters, any highlight

PCR Product: 7456bp

Sequence

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTT
GTTATGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCT
GAAACATGTGCTGTGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTT
GAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGG
CCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCT
CGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGTATAA
GAGGAAGGAATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAATGT
CTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTCAGGGCTGGAGGTGGGACCT
GCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTAATCC
TTTACATTGTCTATGATGCCAAGACCTTTGTTCACGTGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTGA
CCCTGACACATCCCCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACTCAGAGGCTGGCGGG
ATCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTC
TTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGCAACCCACCCCTACATCTGGTG
CCCAACGTGGAGGCTTTTCTCTAGGGTGAAGGTACGTTCGAGCGTGGTCATTGAGGACAAGTCGACGAGAGATCC
CGAGTACGTCTACAGTCAGCCTTACGGTAAGCTTGTGCGCTCGGAAGAAGCTAGGGTGATAATGGGGCAAACTAA
AAGTAAAATTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGGAGTTAAAGT
ATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAACTTT
AGATCTAAAAGACTGGAAAAGAATTGGTAAGGAACTAAAACAAGCAGGTAGGAAGGGTAATATCATTCCACTTAC
AGTATGGAATGATTGGGCCATTATTAAAGCAGCTTTAGAACCATTTCAAACAGAAGAAGATAGTGTTTCAGTTTC
TGATGCCCCTGGAAGCTGTATAATAGATTGTAATGAAAAGACAAGGAAAAAATCCCAGAAAGAAACCGAAAGTTT
ACATTGCGAATATGTAGCAGAGCCAGTAATGGCTCAGTCAACGCAAAATGCTGACTATAATCAATTACAGGAGGT

180

GATATATCCTGAAACGTTAAAATTAGAAGGAAAAGGTCCAGAATTAATGGGGCCATCAGAGTCTAAACCACGAGG
CACAAGTCCTCTTCCAGCAGGTCAGGTGCCCGTAACATTACAACCTCAAAAGCAGGTTAAAGAAAATAAGACCCA
ACCGCCAGTAGCCTACCAATACTGGCCGCCGGCTGAACTTCAGTATCAGCCACCCCCAGAAAGTCAGTATGGATA
TCCAGGAATGCCCCCAGCACCACAGGGCAGGGCGCCATACCCTCAGCCGCCCACTAGGAGACTTAATCCTACGGC
ACCACCTAGTAGACAGGGTAGTGAATTACATGAAATTATTGATAAATCAAGAAAGGAAGGAGATACTGAGGCATG
GCAATTCCCAGTAACGTTAGAACTGATGCCACCTGGAGAAGGAGCCCAAGAGGGAGAGCCTCCCACAGTTGAGGC
CAGATACAAGTCTTTTTCGATAAAAATGCTAAAAGATATGAAAGAGGGAGTAAAACAGTATGGACCCAACTCCCC
TTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCCTTATGATTGGGAGATTCTGGCAAA
ATCGTCTCTCTCACCCTCTCAATTTTTACAATTTAAGACTTGGTGGATTGATGGGGTACAAGAACAGGTCCGAAG
AAATAGGGCTGCCAATCCTCCAGTTAACATAGATGCAGATCAACTATTAGGAATAGGTCAAAATTGGAGTACTAT
TAGTCAACAAGCATTAATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAAT
CCAAGACCCAGGAAGTACCTGCCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCCCTATCCTGATTTTGT
GGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCCGATGAAAAAGCCCGTAAGGTCATAGTGGAGTTAATGGC
ATATGAAAACGCCAATCCTGAGTGTCAATCAGCCATTAAGCCATTAAAAGGAAAGGTTCCTGCAGGATCAGATGT
AATCTCAGAATATGTAAAAGCCTGTGATGGAATGGGAGGAGCTATGCATAAAGCTATGCTTATGGCTCAAGCAAT
AACAGGAGTTGTTTTAGGAGGACAAGTTAGAACATTTGGAGGAAAATGTTATAATTGTGGTCAAATTGGTCACTT
AAAAAAGAATTGCCCAGTCTTAAACAAACAGAATATAACTATTCAAGCAACTACAACAGGTAGAGAGCCACCTGA
CTTATGTCCAAGATGTAAAAAAGGAAAACATTGGGCTAGTCAATGTCGTTCTAAATTTGATAAAAATGGGCAACC
ATTGTCGGGAAACGAGCAAAGGGGCCAGCCTCAGGCCCCACAACAAACTGGGGCATTCCCAATTCAGCCATTTGT
TCCTCAGGGTTTTCAGGGACAACAACCCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCAGTTACCACAATACAA
CAATTGTCCCCCGCCACAAGCGGCAGTGCAGCAGTAGATTTATGTACTATACAAGCAGTCTCTCTGCTTCCAGGG
GAGCCCCCACAAAAAATCCCTACAGGTGTATATGGCCCCCTGCCTGAGGGGACTGTAGGACTAATCTTGGGAAGA
TCAAGTCTAAATCTAAAAGGAGTTCAAATTCATACTAGTGTGGTTGATTCAGACTATAAAGGCGAAATTCAATTG
GTTATTAGCTCTTCAATTCCTTGGAGTGCCAGTCCAGGAGACAGGATTGCTCAATTATTACTCCTGCCATATATT
AAGGGTGGAAATAGTGAAATAAAAAGAACAGGAGGGCTTGGAAGCACTGATCCGACAGGAAAGGCTGCATATTGG
GCAAGTCAGGTCTCAGAGAACAGACCTGTGTGTAAGGCCATTATTCAAGGAAAACAGTTTGAAGGGTTGGTAGAC
ACTGGAGCAGATGTCTCTATCATTGCTTTAAATCAGTGGCCAAAAAATTGGCCTAAACAAAAGGCTGTTACAGGA
CTTGTCGGCATAAGCACAGCCTCAGAAGTGTATCAAAGTACGGAGATTTTACATTGCTTAGGGCCAGATAATCAA
GAAAGTACTGTTCAGCCAATGATTACTTCAATTCCTCTTAATCTGTGGGGTCGAGATTTATTACAACAATGGGGT
GTGGAAATCACCATGCCCGCTCCATTATATAGCCCCACGAGTCAAAAAATCATGACCAAGATGGGATATATACCA
GGAAAGGGACTAGGGAAAAATGAAGATGGCATTAAAGTTCCAGTTGAGGCTAAAATAAATCAAGAAAGAGAAGGA
ATAGGGTATCCTTTTTAGGGGCGGCCACTGTAGAGCCTCCTAAACCCATACCATTAACTTGGAAAACAGAAAAAC
CAGTGTGGGTAAATCAGTGGCCGCTACCAAAACAAAAACTGGAGGCTTTACATTTATTAGCAAATGAACAGTTAG
AAAAGGGTCACATTGAACCTTCGTTCTCACCTTGGAATTCTCCTGTGTTTGTAATTCAGAAGAAATCAGGCAAAT
GGCGTATGTTAACTGACTTAAGGGCTGTAAACGCCGTAATTCAACCCATGGGGCCTCTCCAACCTGGGTTGCCCT
CTCCAGCCATGATCCCAAAAGATTGGCCTTTAATTATAATTGATCTAAAGGATTGCTTTTTTACCATCCCTCTGG
CAGAGCAGGATTGTGAAAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGAACCAGCCACCAGGTTTCAGT
GGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATTTGTCAGACTTTTGTAGGTCGAGCTCTTCAACCAG
TTAGAGACAAGTTTTCAGACTGTTATATCATTCATTATATTGATGATATTTTATGTGCTGCAGAAACGAAAGATA
AATTAATTGACTGTTATACATTTCTGCAAGCAGAGGTTGCCAATGCTGGACTGGCAATAGCATCTGATAAGATCC
AAACCTCTACTCCTTTTCATTATTTAGGGATGCAGATAGAAAATAGAAAAATTAAGCCACAAAAAATAGAAATAA
GAAAAGACACATTAAAAACACTAAATGATTTTCAAAAATTACTAGGAGATATTAATTGGATTCGGCCAACTCTAG
GCATTCCTACTTATGTCATGTCAAATTTGTTCTCTATCTTAAGAGGAGACTCAGACTTAAATAGTAAAGAATGT
TAACCCCAGAGACAACAAAAGAAATTAAATTAGTGGAAGAAAAAATTCAGTCAGCACAAATAAATAGAATAGATC

181

```
CCTTAGCCCCACTCCGACTTTTGATTTTTGCCACTGCACATTCTCCAATAGGCATCATTATTCAAAATACTGATC
TTGTGGAGTGGTCATTCCTTCCTCACAGTACAGTTAAGACTTTTACATTGTACTTGGATCAAATAGCTACATTAA
TCGGTCAGACAAGATTACGAATAATAAAATTATGTGGGAATGACCCAGACAAAATAGTTGTCCCTTTAACCAAGG
AACAAGTTAGACAAGCCTTTATCAATTCTGGTGCATGGCAGATTGGTCTTGCTAATTTTGTGGGAATTATTGATA
ACCATTACCCAAAAACAAAGATCTTCCAGTTCTTAAAATTGACTACTTGGATTCTACCTAAAATTACCAGACGTG
AACCTTTAGAAAATGCTCTAACAGTATTTACTGATGGTTCCAGCAATGGAAAAGCAGCTTACACAGGGCTGAAAG
AACGAGTAATCAAAACTCCATATCAATCGGCTCAAAGAGCAGAGTTGGTTGCAGTCATTACAGTGTTACAAGATT
TTGACCAGCCTATCAATATTATATCAGATTCTGCATATGTAGTACAGGCTACAAGGGATGTTGAGACAGCTCTAA
TTAAATATAGCATGGATGATCAGTTAAACCAGCTATTCAATTTATTACAACAAACTGTAAGAAAAAGAAATTTCC
CATTTTATATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTGACTAAAGCAAATGAACAAGCTGACT
TACTGGTATCATCTGCACTCATAAAAGCACAAGAACTTCATGCTTTGACTCATGTAAATGCAGCAGGATTAAAAA
ACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACATTGCACCCAGTGTCAAGTCTTACACCTGC
CCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCATTATGGCAAATGGATGTCACACATGTAC
CTTCATTTGGAAGATTATCATATGTTCATGTAACAGTTGATACTTATTCACATTTCATATGGGCAACTTGCCAAA
CAGGAGAAAGTACTTCCCATGTTAAAAAACATTTATTGTCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCA
AAACTGACAATGGACCAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAA
CAGGAATTCCTTATAATTCCCAAGGACAGGCCATAGTTGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTA
AACAAAAAGAAGGGGGAGACAGTAAGGAGTGTACCACTCCTCAGATGCAACTTAATCTAGCACCCTATACTTTAA
ATTTTTTAAACATTTATAGAAATCAGACTACTACTTCTGCAGAACAACATCTTACTGGTAAAAAGAACAGCCCAC
ATGAAGGAAAACTAATTTGGTGGAAAGATAATAAAAATAAGACATGGGAAATAGGGAAGGTGATAACGTGGGGGA
GAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAGCTTCCTGTTTGGATGCCCACTAGACATTTGAAGTTCTACA
ATGAACCCATCGGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACACCGCAATCGAGCACCGTTGACTCAC
AAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCGACTTGG
GCACAACTAAAGAAGCTGACGCAGTTAGCTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCCAGAGAGT
ATGCTGCTTGCAGCCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTAAC
TATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGGATAATCCTATAGAAATA
TATGTTAATGATAGTGTATGGGTACCTGGCCCCACAGATGATTGCTGCCCTGCCAAACCTGAGGAAGAAGGGATG
ATGATAAATATTTCCATTGGGTATCGTTATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAATGCCTGCA
GTCCAAAATTGGTTGGTAGAAGTACCTACTGTCAGTCCCATCAGTAGATTCACTTATCACATGGTAAGCGGGATG
TCACTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTAAAATTTAGACCTAAAGGGAAA
CCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAGTTTTAGTTTGGGAAGAATGTGTGGCCAATAGT
GCGGTGATATTACAAAACAATGAATTCGGAACTCTTATAGATTGGGCACCTCGAGGTCAATTCTACCACAATTGC
TCAGGACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTTTAGAC
AAACATAAGCATAAAAAATTGCAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATCTCTACCGCAAGACCA
AAAATAATAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTACTGTGGCCTCACACCACATTAGA
ATTTGGTCTGGAAATCAAACTTTAGAAACAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCCAGTCTA
ACAGTTCCTTTACAAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCC
CAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATTGGCAACACCGTATTCTG
CTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGGACCGACCGTGGGAGGCCTCACCATCCGTCCAT
ATTTTGACTGAAGTATTAAAAGGTGTTTTAAATAGATCCAAAAGATTCATTTTTACTTTAATTGCAGTGATTATG
GGATTAATTGCAGTCACAGCTACGGCTGCTGTAGCAGGAGTTGCATTGCACTCTTCTGTTCAGTCAGTAAACTTT
GTTAATGATTGGCAAAATAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATCAAAAATTGGCAAATCAA
ATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGGCTCATGAGCTTAGAACATCGTTTCCAGTTACAA
TGTGACTGGAATACGTCAGATTTTTGTATTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTT
```

```
AGATGCCATCTACAGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAAGAACAAATTTTTGAAGCA
TCAAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAATTGCAGGAGTTGCTGATGGCCTCGCAAATCTTAAC
ACTGTCACTTGGGTTAAGACCATTGGAAGTACTACAATTATAAATCTCATATTAATCCTTGTGTGCCTGTTTTGT
CTGTTGTTAGTCTACAGGTGTACCCAACAGCTCCGACGAGACAGCGACCATCGAGAACGGGCCATGATGACGATG
GTGGTTTTGTCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAG
AAAGAAGTAGACATAGGAGACTCCATTTTGTTATGTATTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCT
ATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAACTCAGAGTTGAATGGATTAAGGGCGGT
GCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCCCTAATCTCAA
GTACCCAGGGACACAAAAACTGCAGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTC
TCCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCCGTAA
AGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGAATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGT
CTCCTCCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAA
AAACCGCCTCAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGTGTA
TGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTTTGTTCACGTGTTTGTCTG
CTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACACCCACAGATGATCAATA
AATACTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTT
TCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGT
GTAGGGGCAACCCACCCCTACA
```

AP4.2 LTR HERV Sequence Primers

LTR Regions

LTR Primer Pair 1
ATTAAGGGCGGTGCAGGATG
TGATCATCTGTGGGTGTTTCTC

Forward Primer
Reverse Primer

PCR Product: 8765bp

LTR Primer Pair 2
GATCAGATTGTTACTGTGTC
AACCAGCGTTCAGCATATGG

Forward Primer
Reverse Primer

PCR Product: 8514bp

LTR Primer Pair 3
GAAGGCAGCATGCTCCTTAAG
GAGGATCCCGCCAGCCTCTGAG

Forward Primer
Reverse Primer

PCR Product: 8612bp

Inter Primer Overlaps

LTR Reverse Primer 2 and 3 overlap

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTT
GTTATGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCT
GAAACATGTGCTGTGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTT
GAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGG
CCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCT
CGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGTATAA
GAGGAAGGAATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTCCCTGTCCCTGGGCAATGGAATGT
CTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTCAGGGCTGGAGGTGGGACCT
GCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTAATCC
TTTACATTGTCTATGATGCCAAGACCTTTGTTCACGTGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTGA
CCCTGACACATCCCCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACTCAGAGGCTGGCGGG
ATCCTCCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTC
TTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCACCCCTACATCTGGTG
CCCAACGTGGAGGCTTTTCTCTAGGGTGAAGGTACGCTCGAGCGTGGTCATTGAGGACAAGTCGACGAGAGATCC
CGAGTACGTCTACAGTCAGCCTTACGGTAAGCTTGTGCGCTCGGAAGAAGCTAGGGTGATAATGGGGCAAACTAA
…
CTGTTGTTAGTCTACAGGTGTACCCAACAGCTCCGACGAGACAGCGACCATCGAGAACGGGCCATGATGACGATG
GTGGTTTTGTCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAG
AAAGAAGTAGACATAGGAGACTCCATTTTGTTATGTATTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCT
ATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAACTCAGAGTTGAATGGATTAAGGGCGGT
GCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCCCTAATCTCAA
GTACCCAGGGACACAAAAACTGCAGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTC
TCCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCCGTAA
AGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGAATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGT
CTCCTCCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAA
AAACCGCCTCAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGTGTA
TGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTTTGTTCACGTGTTTGTCTG
CTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACACCCACAGATGATCAATA
AATACTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTT
TCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGT
GTAGGGGCAACCCACCCCTACA

Appendix 5 – HERV Primer Pair 3 Primer BLAST

Sequence (5'->3')

Forward primer CTCGAGCGTGGTCATTGAG

Reverse primer CTCTCTTGCTTTTCCCCACATT

Total Sequences expected to be amplified: 56

HERV Sequences: 7

Sequences around 7500 bp: 20

HERV Sequences around 7500 bp: 5

7500 bp HERV sequences are highlighted green, while other sized HERV sequences are highlighted light blue.

```
product length = 4414
Features associated with this product:
   choline transporter-like protein 5 isoform b

   choline transporter-like protein 5 isoform x2

Forward primer  1          CTCGAGCGTGGTCATTGAG  19
Template        75378097   ...................  75378115

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        75382510   ......................  75382489


product length = 7271
Features associated with this product:
   cd48 antigen isoform x2

   cd48 antigen isoform 1 precursor

Forward primer  1          CTCGAGCGTGGTCATTGAG  19
Template        160691797  ...................  160691815

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        160699067  ......................  160699046


product length = 7221
Features flanking this product:
   13342 bp at 5' side: protein misato homolog 1 isoform 12
   24822 bp at 3' side: yy1-associated protein 1 isoform 5
```

```
Forward primer   1          CTCGAGCGTGGTCATTGAG   19
Template         155634834  ...................   155634816


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         155627614  .....................   155627635


product length = 7360
Features flanking this product:
   134053 bp at 5' side: monocarboxylate transporter 1
   2687 bp at 3' side: leucine-rich repeats and immunoglobulin-like domains
prot...


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         113070719  T....AAT...........C..   113070698


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         113063360  T.....AT..............   113063381
```

>NC_000003.12 Homo sapiens chromosome 3, GRCh38.p14 Primary Assembly

```
product length = 7162
Features associated with this product:
   endogenous retrovirus group k member 5 gag polyprotein is...

   endogenous retrovirus group k member 5 gag polyprotein is...


Forward primer   1          CTCGAGCGTGGTCATTGAG   19
Template         101692905  ...................   101692923


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         101700066  .....................   101700045


product length = 7508
Features flanking this product:
   43036 bp at 5' side: translation initiation factor if-2-like
   73242 bp at 3' side: ropporin-1b


Forward primer   1          CTCGAGCGTGGTCATTGAG   19
Template         125891305  ...................   125891323


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         125898812  .....................   125898791


product length = 7216
Features flanking this product:
   5571 bp at 5' side: nucleolus and neural progenitor protein isoform x4
   179122 bp at 3' side: uncharacterized protein loc124909486


Forward primer   1          CTCGAGCGTGGTCATTGAG   19
Template         113032432  ...................   113032414


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         113025217  .....................   113025238


product length = 7222
Features flanking this product:
   28445 bp at 5' side: lipase member h isoform x3
```

```
     15696 bp at 3' side: sentrin-specific protease 2


Forward primer   1          CTCGAGCGTGGTCATTGAG   19
Template         185570717  ..................   185570699


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT   22
Template         185563496  .....................   185563517


product length = 4967
Features flanking this product:
   4636 bp at 5' side: mitochondrial mrna pseudouridine synthase rpusd3
isoform 5
   13542 bp at 3' side: cell death activator cide-3 isoform 3


Forward primer   1          CTCGAGCGTGGTCATTGAG   19
Template         9853591    ......T............   9853573


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT   22
Template         9848625    TGA...CT..............   9848646


product length = 4920
Features flanking this product:
   148435 bp at 5' side: myelin-associated oligodendrocyte basic protein
isoform b
   244583 bp at 3' side: rab effector myrip isoform a


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT   22
Template         39656233   A......T..............   39656212


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT   22
Template         39651314   T...G.AT..............   39651335


product length = 2616
Features flanking this product:
   15030 bp at 5' side: lhfpl tetraspan subfamily member 4 protein isoform
x1
   79260 bp at 3' side: myotubularin-related protein 14 isoform 20


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT   22
Template         9570323    T.....AT......A.......   9570302


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT   22
Template         9567708    T....A.T........TT....   9567729
```

>NC_000007.14 Homo sapiens chromosome 7, GRCh38.p14 Primary Assembly

```
product length = 7513
Features flanking this product:
   314625 bp at 5' side: protein sidekick-1 isoform 2
   965 bp at 3' side: endogenous retrovirus group k member 6 env polyprotein


Forward primer   1          CTCGAGCGTGGTCATTGAG   19
Template         4590886    ..................   4590868


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT   22
Template         4583374    .....................   4583395
```

```
product length = 7512
Features associated with this product:
    endogenous retrovirus group k member 6 env polyprotein

Forward primer    1          CTCGAGCGTGGTCATTGAG   19
Template          4599389    ..................    4599371

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          4591878    .....................   4591899


product length = 558
Features flanking this product:
    46130 bp at 5' side: uncharacterized protein loc112267992
    231081 bp at 3' side: glycoprotein-n-acetylgalactosamine 3-beta-
galactosyltrans...

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          7003238    A...T..CT........T....   7003217

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          7002681    .......T.....T.......C   7002702


product length = 6998
Features flanking this product:
    58174 bp at 5' side: vacuolar fusion protein ccz1 homolog b
    64201 bp at 3' side: uncharacterized protein loc112267992

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          6891367    .CT........G...T.....G   6891346

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          6884370    .......T.....G..T...C.   6884391


product length = 1997
Features flanking this product:
    192354 bp at 5' side: mitochondrial inner membrane protease subunit 2
isoform x10
    12477 bp at 3' side: dedicator of cytokinesis protein 4 isoform x7

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          111715796  ..AAT..C..........T...   111715775

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          111713800  .C.....T..C.C........A   111713821


product length = 2565
Features flanking this product:
    332880 bp at 5' side: ankyrin repeat domain-containing protein 7
    1699228 bp at 3' side: potassium voltage-gated channel subfamily d member
2 isof...

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          118575404  ..GGG..C........T.....   118575383

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          118572840  T.T...ACT.............   118572861
```

>NC_000011.10 Homo sapiens chromosome 11, GRCh38.p14 Primary Assembly

```
product length = 7507
Features associated with this product:
   endogenous retrovirus group k member 19 rec protein isofo...

   endogenous retrovirus group k member 25 env polyprotein i...

Forward primer  1          CTCGAGCGTGGTCATTGAG  19
Template        101696074  ...................  101696092

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        101703580  ......................  101703559


product length = 7202
Features associated with this product:
   endogenous retrovirus group k member 7 env polyprotein-like

   endogenous retrovirus group k member 7 env polyprotein-like

Forward primer  1          CTCGAGCGTGGTCATTGAG  19
Template        118729163  ...A...............  118729145

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        118721962  ......................  118721983
```

>NC_000019.10 Homo sapiens chromosome 19, GRCh38.p14 Primary Assembly

```
product length = 7846
Features flanking this product:
   3510586 bp at 5' side: zinc finger protein 254 isoform x4
   1561137 bp at 3' side: cytochrome b-c1 complex subunit rieske,
mitochondrial

Forward primer  1          CTCGAGCGTGGTCATTGAG  19
Template        27646410   ...................  27646392

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        27638565   ......................  27638586


product length = 4871
Features flanking this product:
   153718 bp at 5' side: zinc finger protein 98
   72588 bp at 3' side: zinc finger protein 492

Forward primer  1          CTCGAGCGTGGTCATTGAG  19
Template        22575941   ...........T......  22575959

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        22580811   .......T........T.....  22580790


product length = 3597
Features associated with this product:
   zinc finger protein 91 isoform x1

   zinc finger protein 91 isoform 2

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
```

```
Template          23369593  T.....AT.............  23369572

Reverse primer  1            CTCTCTTGCTTTTCCCCACATT  22
Template          23365997  T....AAT...........C..  23366018


product length = 2772
Features flanking this product:
   16997 bp at 5' side: zinc finger protein 562 isoform x1
   71927 bp at 3' side: zinc finger protein 846 isoform 10

Reverse primer  1            CTCTCTTGCTTTTCCCCACATT  22
Template          9680511  T....AAT...........C..  9680490

Reverse primer  1            CTCTCTTGCTTTTCCCCACATT  22
Template          9677740  A......T........T...A.  9677761
```

>NC_000022.11 Homo sapiens chromosome 22, GRCh38.p14 Primary Assembly

```
product length = 7220
Features flanking this product:
   8539 bp at 5' side: proline dehydrogenase 1, mitochondrial isoform 2
   82619 bp at 3' side: protein fam246c

Forward primer  1            CTCGAGCGTGGTCATTGAG  19
Template          18939685  ...................  18939703

Reverse primer  1            CTCTCTTGCTTTTCCCCACATT  22
Template          18946904  ......................  18946883
```

>NC_000005.10 Homo sapiens chromosome 5, GRCh38.p14 Primary Assembly

```
product length = 7221
Features associated with this product:
   delta-sarcoglycan isoform 1

   delta-sarcoglycan isoform x3

Forward primer  1             CTCGAGCGTGGTCATTGAG  19
Template          156665874  ...................  156665856

Reverse primer  1             CTCTCTTGCTTTTCCCCACATT  22
Template          156658654  ......................  156658675


product length = 7512
Features flanking this product:
   1559289 bp at 5' side: translation initiation factor if-2
   772369 bp at 3' side: cadherin-6 isoform 1 preproprotein

Forward primer  1            CTCGAGCGTGGTCATTGAG  19
Template          30495104  .......A...........  30495086

Reverse primer  1            CTCTCTTGCTTTTCCCCACATT  22
Template          30487593  .G...................  30487614


product length = 7909
Features flanking this product:
   304893 bp at 5' side: potassium/sodium hyperpolarization-activated cyclic
nucle...
```

```
Forward primer  1          CTCGAGCGTGGTCATTGAG  19
Template        46008893   ...A...A...........  46008875

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        46000985   .......T..............  46001006
```

product length = 1750
Features associated with this product:
   mastermind-like protein 1

```
Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        179763598  ..G.TC.T.............C  179763577

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        179761849  ...C...C.............C  179761870
```

product length = 852
Features associated with this product:
   rbm27-pou4f3

```
Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        146302669  .C..G.GTA.............  146302648

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        146301818  ....G.GTA.............  146301839
```

>NC_000006.12 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly

product length = 7479
Features associated with this product:
   meiosis-specific protein mei4

   meiosis-specific protein mei4

```
Forward primer  1          CTCGAGCGTGGTCATTGAG  19
Template        77725363   ...................  77725345

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        77717885   ......................  77717906
```

>NC_000008.11 Homo sapiens chromosome 8, GRCh38.p14 Primary Assembly

product length = 7512
Features associated with this product:
   beta-defensin 107 precursor

```
Forward primer  1          CTCGAGCGTGGTCATTGAG  19
Template        7506334    ...................  7506316

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        7498823    ......................  7498844
```

product length = 2097
Features flanking this product:
   577683 bp at 5' side: collagen alpha-1(xxii) chain isoform x9
   155307 bp at 3' side: potassium channel subfamily k member 9

```
Forward primer    1          CTCGAGCGTGGTCATTGAG  19
Template          139462950  ..................   139462932


Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          139460854  ....................    139460875


product length = 1887
Features flanking this product:
   21571 bp at 5' side: solute carrier organic anion transporter family
member 5a...
   195947 bp at 3' side: pr domain zinc finger protein 14

Forward primer    1          CTCGAGCGTGGTCATTGAG  19
Template          69854243   A.....GC..........C  69854261

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          69856129   .......T..............  69856108


product length = 6253
Features flanking this product:
   868749 bp at 5' side: yth domain-containing family protein 3 isoform c
   492612 bp at 3' side: uncharacterized protein loc124900252

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          64084706   T.....G.T..C.......C..  64084685

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          64078454   ..T.......C....TT.....  64078475


product length = 2809
Features associated with this product:
   tyrosine-protein kinase lyn isoform a

   tyrosine-protein kinase lyn isoform b

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          56001482   ....G.G...CA.........G  56001461

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          55998674   ........A..AA.T......G  55998695
```

>NC_000010.11 Homo sapiens chromosome 10, GRCh38.p14 Primary Assembly

```
product length = 6139
Features associated with this product:
   atp-binding cassette sub-family c member 2 isoform x3

   atp-binding cassette sub-family c member 2 isoform x2

Forward primer    1          CTCGAGCGTGGTCATTGAG  19
Template          99827916   ..................   99827898

Reverse primer    1          CTCTCTTGCTTTTCCCCACATT  22
Template          99821778   ....................    99821799


product length = 7504
```

```
Features associated with this product:
   endogenous retrovirus group k member 6 env polyprotein


Forward primer   1          CTCGAGCGTGGTCATTGAG   19
Template         6832630    .......A...........   6832612


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         6825127    ..G....T..............  6825148


product length = 2193
Features flanking this product:
   531409 bp at 5' side: uncharacterized protein loc107983989
   31155 bp at 3' side: wiskott-aldrich syndrome protein family member 1-
like


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         3035547    .........C.GC.G......A  3035526


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         3033355    G.......T..G.AG.......  3033376
```

>NC_000012.12 Homo sapiens chromosome 12, GRCh38.p14 Primary Assembly

```
product length = 7498
Features associated with this product:
   endogenous retrovirus group k member 21 env polyprotein

   endogenous retrovirus group k member 21 env polyprotein


Forward primer   1          CTCGAGCGTGGTCATTGAG   19
Template         58335904   ...................   58335886


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         58328407   .....................  58328428


product length = 73
Features flanking this product:
   301522 bp at 5' side: transcription factor sox-5 isoform x16
   452566 bp at 3' side: uncharacterized protein loc124902897


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         24252117   A.G....T........AC....  24252096


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         24252045   .AT.T........G..A.....  24252066
```

>NC_000004.12 Homo sapiens chromosome 4, GRCh38.p14 Primary Assembly

```
product length = 5280
Features flanking this product:
   26692 bp at 5' side: putative tripartite motif-containing protein 61
isoform 1
   38100 bp at 3' side: tripartite motif-containing protein 60


Forward primer   1          CTCGAGCGTGGTCATTGAG   19
Template         164996693  ......T............   164996711


Reverse primer   1          CTCTCTTGCTTTTCCCCACATT  22
Template         165001972  A......T..............  165001951
```

```
product length = 6812
Features associated with this product:
   calcium uniporter regulatory subunit mcub, mitochondrial

   calcium uniporter regulatory subunit mcub, mitochondrial ...

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        109603973  .......T..............  109603952

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        109597162  T....AAT...........C..  109597183


product length = 1790
Features flanking this product:
   14678 bp at 5' side: amyloid beta precursor protein binding family b
member 2 ...
   207256 bp at 3' side: ubiquitin carboxyl-terminal hydrolase isozyme l1

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        41049720   T...G.AT..............  41049699

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        41047931   .....C.-..............  41047951


product length = 7165
Features flanking this product:
   138322 bp at 5' side: alpha-2c adrenergic receptor
   275322 bp at 3' side: proton channel otop1

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        3913480    .CT........G...T.....G  3913459

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        3906316    .......T.....G..T...C.  3906337


product length = 6986
Features flanking this product:
   15511 bp at 5' side: zinc finger protein 141 isoform 2
   45337 bp at 3' side: zinc finger protein 721

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        396357     T....AAT...........C..  396336

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        389372     .AG..C.A...........G..  389393
```

>NC_000002.12 Homo sapiens chromosome 2, GRCh38.p14 Primary Assembly

```
product length = 453
Features associated with this product:
   uncharacterized protein loc124907920

Forward primer  1          CTCGAGCGTGGTCATTGAG  19
Template        186521096  .......A..A........  186521114

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
```

```
Template        186521548  A......T.............  186521527
```

product length = 6198
Features associated with this product:
   dystrobrevin beta isoform 28

   dystrobrevin beta isoform 12

```
Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        25540628   T....AGT...........C..  25540607

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        25534431   T...G..T..............  25534452
```

product length = 1663
Features flanking this product:
   8019 bp at 5' side: translation initiation factor if-2, mitochondrial
isoform x2
   18172 bp at 3' side: girdin isoform 3

```
Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        55273538   ..T............TA..C..  55273517

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        55271876   AC.....TT..........T..  55271897
```

product length = 7104
Features flanking this product:
   318763 bp at 5' side: spermatid nuclear transition protein 1
   618559 bp at 3' side: tensin-1 isoform 3

```
Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        217185899  AA..T..AT.............  217185878

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        217178796  T.......A..C..T.T.....  217178817
```

product length = 1673
Features flanking this product:
   70573 bp at 5' side: cdc42 effector protein 3
   210446 bp at 3' side: regulator of microtubule dynamics protein 2 isoform
x6

```
Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        37718831   G.......G..C..T....T..  37718810

Reverse primer  1          CTCTCTTGCTTTTCCCCACATT  22
Template        37717159   ..G.A...T...C........G  37717180
```

>NC_000023.11 Homo sapiens chromosome X, GRCh38.p14 Primary Assembly

product length = 6823
Features associated with this product:
   sperm protein associated with the nucleus on the x chromo...

   sperm protein associated with the nucleus on the x chromo...

```
Forward primer   1              CTCGAGCGTGGTCATTGAG   19
Template         141583812      ....G...........C..   141583830


Forward primer   1              CTCGAGCGTGGTCATTGAG   19
Template         141590634      ....G...........C..   141590616


product length = 2067
Features associated with this product:
   centromere protein v-like protein 3

Forward primer   1              CTCGAGCGTGGTCATTGAG   19
Template         51618181       GCGC...........C...   51618163


Reverse primer   1              CTCTCTTGCTTTTCCCCACATT   22
Template         51616115       AC.........C...T.T...   51616136


product length = 3433
Features flanking this product:
   448223 bp at 5' side: homeobox protein arx
   671594 bp at 3' side: melanoma-associated antigen b18


Reverse primer   1              CTCTCTTGCTTTTCCCCACATT   22
Template         25467391       T.T.T.G......G........   25467370


Reverse primer   1              CTCTCTTGCTTTTCCCCACATT   22
Template         25463959       T.........GC..T...G...   25463980
```

>NC_000020.11 Homo sapiens chromosome 20, GRCh38.p14 Primary Assembly

```
product length = 4306
Features associated with this product:
   receptor-type tyrosine-protein phosphatase alpha isoform ...

   receptor-type tyrosine-protein phosphatase alpha isoform ...

Reverse primer   1              CTCTCTTGCTTTTCCCCACATT   22
Template         3022959        ..A....T....C.......A.   3022938


Reverse primer   1              CTCTCTTGCTTTTCCCCACATT   22
Template         3018654        T....AAT...........C..   3018675
```

>NC_000016.10 Homo sapiens chromosome 16, GRCh38.p14 Primary Assembly

```
product length = 2801
Features associated with this product:
   cklf-like marvel transmembrane domain-containing protein ...

   cklf-like marvel transmembrane domain-containing protein ...

Reverse primer   1              CTCTCTTGCTTTTCCCCACATT   22
Template         66631517       T....AGT...........C..   66631496


Reverse primer   1              CTCTCTTGCTTTTCCCCACATT   22
Template         66628717       .......TT...C...T....C   66628738
```

>NC_000017.11 Homo sapiens chromosome 17, GRCh38.p14 Primary Assembly

```
product length = 3512
```

```
Features flanking this product:
    49949 bp at 5' side: proton-coupled folate transporter isoform x2
    14361 bp at 3' side: solute carrier family 13 member 2 isoform x5


Reverse primer  1         CTCTCTTGCTTTTCCCCACATT  22
Template        28459351  T....AAT...........C..  28459330


Reverse primer  1         CTCTCTTGCTTTTCCCCACATT  22
Template        28455840  T....AAT...........C..  28455861
```