Imperial College London

RESCUE OF MURINE GAMMARETROVIRUSES IN HUMAN CELLS

Division of Infectious Diseases
Department of Medicine

Submitted for the degree of
Doctor of Philosophy

Ourania K. Kakisi, MBBS

2018

THIS PAGE INTENTIONALLY LEFT BLANK

ACRONYMS AND ABBREVIATIONS

bp	Base pairs
BSA	Bovine Serum Albumin
CA	Capsid protein
cDNA	Complementary DNA
Cfu	Colony forming unit
DMEM	Dulbecco's Modified Eagle Medium
dNTPs	Deoxynucleotide Triphosphates
DTT	Dithiothreitol
EDTA	Ethylene-diamine-tetra-acetic Acid
EGF	Epidermal Growth Factor
FCS	Fetal Calf Serum
GADPH	Glyseraldehyde-3-phosphate dehydrogenase
GALV	Gibbon-Ape Leukaemia Virus
HPRT	Hypoxanthine-guanine phospho-ribosyl-transferase
iGFP	Inducible Green Fluorescent Protein
IN	Integrase protein
Kbp	kilo basepair
kDA	kilo-Daltons
LTR	Long Terminal Repeat
MEF	Mouse Embryonic Fibroblast
MLV	Murine Leukaemia Virus
mRNA	Messenger Ribonucleic Acid
NC	Nucleocapsid
NHSBT	National Health Service Blood Transfusion
NIH	National Institutes of Health
PBS	Phosphate Buffer Saline
PCa	Prostate Cancer
PERV	Porcine Endogenous Retrovirus
PHE	Public Health England
POL	Polymerase protein
RTase	Reverse transcriptase
STDEV	Standard Deviation
tRNA	Transfer RNA
XMRV	Xenotropic murine leukaemia virus- related retrovirus

ABSTRACT

Murine leukaemia viruses (MLVs) are simple γ-*retroviruses* of mice that are endogenous to the mouse genome, with several being replication-competent. In 2006, a research group using virus arrays, identified a novel MLV recombinant, XMRV, in human prostate cancer (PCa) tissue and cell lines. The virus appeared to infect and replicate efficiently in human cells. Disease associations with PCa and chronic fatigue syndrome caused a worldwide concern over the presence of XMRV in the human blood supply and patient populations. The final conclusions settled in 2011, and XMRV was widely accepted as a result of contamination in patient samples while its rescue was proven directly due to the propagation of the human cancer cell lines in nude mice *in vivo*.

The re-emerged perception of the ability of MLVs to recombine and form novel recombinants that can infect human cells propelled the need of investigation in all cases where mouse and human cells are placed in contact for biotechnology applications *in vitro* ,such as for the expansion of human epidermal primary stem cells prior human transplantation.

We first sought to screen a variety of mouse embryonic feeder (MEF) cells that are used in human stem cell culture, for the presence of replication-competent retroviruses. To this aim, we developed a tool case of molecular assays to detect MLVs and their products and then tested various MEFs with known endogenous virus activators. We identified one stably expressed ecotropic MLV retrovirus in a CF1-MEF cell line and characterised it with molecular and infectivity assays, as well as next generation sequencing (NGS). Following, we co-cultivated human corneal stem cells with the CF1-MEF cells to test the ability of the CF1-related virus to cause human cell infection and used integration splinkerette PCR and NGS to detect possible integration site sequences in the human stem cells.

In all, this study provides with a comprehensive up to date assessment over the risk of transmission of MLV viruses onto human stem cells for conditions that are comparable to the protocols used in human eye cell transplantation. We conclude that the infection of stem cells is confirmed possible by xenotropic MLVs as demonstrated for limbal eye

stem cells and XMRV, but unlikely as expected for the ecotropic CF1 cell-derived virus; the possibility of recombination in mouse feeders however, cannot be excluded during prolonged co-cultivations.

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS 3
ABSTRACT 4
TABLE OF CONTENTS 6
STATEMENT OF ORIGINALITY 10
AKNOWLEDGEMENTS 11
Chapter 1 15
Introduction 15
1.1 General introduction 16
1.2 Retroviruses 17
1.3 Retroviral virion morphology and classification 22
1.4 Murine Leukaemia Viruses (MLVs) 27
1.4.1 MLV structure and replication 27
1.4.2. Classification of MLVs 29
1.4.2.1 Endogenous and exogenous MLVs 29
1.4.2.2 Classification of MLVs according to receptor usage 29
1.4.3 Recombination in MLVs 34
1.5 The emergence of XMRV 37
1.6 XMRV and the contamination origin 42
1.6.1 XMRV-related contamination 42
1.6.2 Disease dis-association 43
1.6.3 Mouse DNA and MLV contamination 44
1.7 Human stem cell co-cultivation with mouse cells 45
1.7.1 The principle of human and mouse cell co-cultivation 45
1.7.2 MLV virus infectivity for human stem cells 48
1.7.3 Clinical considerations over mouse and human cell co-cultivation 51
1.7.4 The CF1 feeder cell-line 53
1.7.5 CF1 mice and MLV virus expression 56
1.8 Thesis aims and outline 58
Chapter 2 59
Materials and Methods 59
2.1 Cell lines and maintenance 60
2.2 Human stem cell and mouse feeder cell co-cultivation 63
2.3 Chemical stimulation of endogenous retroviruses in mouse cells 64
2.4 Infectivity assays 65
2.4.1 Syncytium assay on NIH-3T3 cells 65
2.4.2 XC plaque forming assay 65
2.5 Molecular Methods 66

2.5.1 Nucleic acid extraction from cell culture supernatant 66
2.5.2 Reverse transcription 67
2.5.3 PCR amplification for MLV gene targets 67
2.5.4. Oligonucleotides designed for PCR amplification 70
2.5.5 Gel electrophoresis, gel extraction and purification 72
2.5.6 Direct (Sanger) sequencing 72
2.5.7 Product-enhanced reverse transcriptase activity assay 73
2.5.8 Transcription PCR 76
2.5.9 Integration site mapping via splinkerette PCR 76
2.5.10 Next generation sequencing 81
2.6 Imaging 81
2.6.1 Immunofluorescence 81
2.6.2 Immunofluorescence using the indicator fluorescent DERSE cell line 82
2.6.3 Transmission electron microscopy 83
Chapter 3 85
Investigation of MLV virus activation in murine cell lines 85
3.1 Introduction 86
3.1.1 5-Iododeoxyuridine (5-IDU) 87
3.1.2 5- Azacytidine- C (5-AZAC) 88
3.2 Pre-assessment of experimental conditions 89
3.2.1 Growth curve characteristics of MEF cells 89
3.2.2 Drug toxicity and effect on cell proliferation of MEF cells 92
3.2.3 Screening PERT assays for candidate chemical stimulants 96
3.3 Induction experiment 98
3.3.1 Assessment of the induction of endogenous MLVs in the established cell line Balb-K 99
3.3.2 Confirmation of molecular sequences involved in the Balb-K induction 104
3.3.3 Assessment of the induction of endogenous MLVs in J2-3T3 cells 107
3.3.4 Assessment of the induction of endogenous MLVs in CF1 cells 114
3.4 Discussion 118
Chapter 4 121
Molecular and virological characterisation of the CF1 cell line-related MLV 121
4.1 Introduction 122
4.2 RESULTS 122
4.2.1 Identification of CF1 related MLV-RNA using MLV gag (Nested) PCR 122
4.2.2 Further identification of the CF1-related MLV-RNA via MLV env PCR 123
4.2.3 Phylogenetic analysis of env gene derived sequences 128
4.2.4 Immunofluorescence staining for MLV Gag in CF1 cells 130
4.2.5 Distinction of RFB-MLV and AKV-MLV via tandem repeat LTR PCR 132
4.2.6 Syncytia formation assays on NIH-3T3 cells 136
4.2.7 Direct infectivity assays by syncytia formation on XC cells 140

4.2.8 Spliced RFB-MLV envelope mRNA is transcribed in CF1 cells 142
4.2.9. Control for Mitomycin C treatment effect on MLV transcription 146
4.2.10 CF1-related MLV virus is unable to infect human DERSE iGFP cells 147
4.2.11 Transmission Electron Microscopy (TEM) 150
4.2.12 Determination of the genomic content of CF1 cells in xenotropic MLV gag 153
4.3 Discussion 155
Chapter 5 159
Full length sequencing of the CF1 -related MLV genome by RNA-Seq 159
5.1 Introduction 160
5.2 Methods 160
5.3 Results 161
5.3.1 Construction of libraries and assembly of full length viral RNA 161
5.3.2 Detailed organisation of the CF1 cell line-related full length consensus sequence 170
5.3.2.1 Coding regions and open reading frames 171
5.3.2.2. Upstream non-coding regions 172
5.4 Variant determination 173
5.5 Alignment of the full-length CF1 MLV consensus 175
5.6 Discussion 179
Chapter 6 181
Viral integration analysis 181
6.1 Introduction 182
6.2 Integration site analysis via splinkerette PCR for CF1-related MLV in genomic DNA 182
6.2.1. Retroviral integration in genomic DNA from cell lines 182
6.2.2. Retroviral integration in genomic DNA post in vitro infection 190
6.3 Next generation sequencing of XMRV integration sites in 22rv1 cells 196
6.3.1 Construction of library and analysis of the results 196
6.3.2 Results 198
6.4 CF1 cell-related MLV integration in CF1 cells by NGS 205
6.5 Discussion 207
Chapter 7 209
CF1-MEF and human corneal stem cell co-cultivation 209
7.1 Introduction 210
7.2 In vitro expansion protocol for limbal stem cells 210
7.3 Detection of MLV in the co-cultivation of human stem and mouse feeder layers 212
7.4 Immunofluorescence staining on co-cultures with anti-MLV Gag 215
7.5 CF1-MLV gag and env mRNA expression in total cell RNA from co-cultivations 219
7.6 Investigation of CF1-MLV integration in human stem cells grown on CF1 feeder layers via splinkerette and next generation sequencing 221
7.6.1 NGS strategy and first iteration mapping to reference 221
7.6.2 Integration site analysis of co-cultivated stem cells 225
7.6.3 Verification of integration sites defined by NGS 233

7.7 Mus dunni infection assay and reverse transcriptase activity for polytropic viruses 237
7.8 CF1 cell-related MLV integration site in genomic DNA inferred from NGS of human stem cell co-cultivations 240
7.9 Discussion 244
Chapter 8 246
DISCUSSION 246
REFERENCES 254
APPENDIX 273

STATEMENT OF ORIGINALITY

I confirm that all work presented in this thesis is my own with the following exceptions. TEM was performed by Matthew Hannah, PHE, London. Gel excision purification and quality control for RNA-Seq NGS sequencing was performed by Dr. Steve Kaye, Imperial College, London. The use of all material from other sources has been properly and fully acknowledged.

A significant part of the Introduction is derived from my co-authored invited publication "The Rise and Fall of XMRV" in the journal Transfusion Medicine, 2013 for which copyright permissions have been sought and gained for reproduction in this thesis and found at the end of the Appendix (12).

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives license. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work.

Ourania K. Kakisi

AKNOWLEDGEMENTS

First of all, I would like to thank my supervisors Professor Myra McClure and Professor Richard S. Tedder who allowed me to start a PhD whilst a clinical research fellow and hence a hospital Microbiology consultant and have always been a source of challenging advice, support and guidance. I have had a sincerely enjoyable and robust learning experience at Imperial and PHE Colindale.

I would also like to acknowledge the warm support of all the people working in the Jefferiss Trust Wing laboratory, who continuously offered technical advice and mainly Dr Steve Kaye, to whom I owe a lot for his exemplary professionalism, his continuous support and his insightful instructions and dr Jeremy Garson for sharing his expertise. For part of this thesis, work was also carried out in Public Health England and the Virus Reference Department, with whom I collaborated for my clinical research fellowship in Transfusion Microbiology. I would also like to extend my thanks to them and especially the Blood Borne Viruses Unit for their support and their friendship. I am also grateful to the NHS Blood and Transplant for their funding and wish to applaud them for a service that goes beyond personal sacrifice for a fundamental duty of care such is transfusion and transplantation.

My parents have been instrumental in maintaining my spirit throughout my education since my first day of school and I wish to thank them for their patience. Their dedication is something I am always grateful. During this thesis preparation, my father passed away always hoping that I have the strength to accomplish my goals and my love and ultimate dedication go to him.

Ourania

LIST OF FIGURES

Figure 1.1 Genomic organisation of Murine Leukaemia viruses 20
Figure 1.3 Model of the mature retroviral virion structure.....23
Figure 1.4.1 Replication cycle of MLVs.....28
Figure 1.4.2 Structures of the Surface Envelope MLV (SU Env-MLV) protein that determine virus tropism 32
Figure 1.7.1 Stem cell co-cultivation with mouse feeder layers: the principle 47
Figure 1.7.4 The origin and current market supply chain of the CF1 stock.....55
Figure 2.5.3 A quantification curve in Real Time RT-PCR for MLV gag 69
Figure 2.5.5 Position of MLV gene-directed primers on alignment of MLV isolates 71
Figure 2.5.7: Product-Enhanced Reverse Transcriptase (PERT) Assay 75
Figure 2.5.9: The splinkerette reaction principle.....79
Figure 2.6 The DERSE reporter-cell circular plasmid 84
Figure 3.2.1 Growth curve determination for the prototype NIH-3T3 cell line.....91
Figure 3.2.2 Cell viability assays on NIH-3T3 cells with varying concentrations of trial inducing agents 93
Figure 3.2.3 Evaluation of selected chemicals on MEF cell lines for the induction of RT activity over 21 days 97
Figure 3.3.1: RT activity and MLV viral load in Balb-K cells 101
Figure 3.3.2 Envelope sequencing PCRs of BALB-K induced supernatants.....105
Figure 3.3.3.1 RT activity and MLV viral load in J2 cells treated with 5-IDU 109
Figure 3.3.3.2 MLV env gene sequencing for J2 cell supernatant 112
Figure 3.3.4 Induction of CF1 cells with 5-IDU and 5-AzaC: results of PERT and RT-PCR 115
Figure 4.2.1 Amplification of MLV gag and env from CF1 cell supernatant 126
Figure 4.2.3 Phylogenetic tree of env sequence consensus from CF1 cell supernatant .129
Figure 4.2.4 Immunofluorescent staining of cells containing MLV Gag 131
Figure 4.2.6 Syncytia formation assay for infectivity assessment on NIH-3T3 cells 139
Figure 4.2.7 XC cell syncytial formation assay for CF1 cells.....141
Figure 4.2.8 Transcription of spliced messenger env RNA from CF1 cells.....144
Figure 4.2.9 MLV gag RNA production in Mitomycin C treated CF1 cells 146
Figure 4.2.10.1 Reporter DERSE cells post-infection with cell supernatant 148
Figure 4.2.10.2 The visualisation of MLV infection of DERSE cells prior iGFP emergence with anti-MLV Gag.....149
Figure 4.2.11 Transmission Electron Microscopy of human and mouse cell lines 151
Figure 4.2.12 Xenotropic/polytropic MLV gag real time PCR results on CF1 cells.....154
Figure 4.2.12 Xenotropic/polytropic MLV gag real time PCR results on CF1 cells 154
Figure 5.3.1 Quality control electrophoresis of CF1 derived library pools.....162
Figure 5.3.1.2 Assembly results from paired-end reads of next generation RNA sequencing on CF1 derived pools 166

Figure 5.5 Mutation analysis of CF1 cell-related MLV....177
Figure 6.2.1.1 Positive MoMLV virus integration control YAC-I....183
Figure 6.2.1.2 XMRV integration site in 22rv1 gDNA....185
Figure 6.2.1.3 The integration sequence obtained via direct sequencing for XMRV in 22rv1 gDNA....186
Figure 6.2.1.4 CF1-related MLV integration site in CF1 gDNA....188
Figure 6.2.1.5 The integration sequence obtained via direct sequencing for CF1-related MLV in CF1 gDNA....189
Figure 6.2.2 CF1-related MLV integration in infected mouse cell line gDNA....191
Figure 6.2.3 The assessment of mouse DNA contamination in the integration PCRs....194
Figure 6.3.1 The analysis algorithm for the confirmation of NGS sequences....197
Figure 6.3.2 Integration sites of XMRV in 22rv1 human prostate cancer cells and their chromosomal locations....201
Figure 6.3.3 Characteristics of XMRV integration sites in 22rv1 cells....203
Figure 6.4: Mouse chromosomal integrations for CF1-MLV in CF1 cells via NGS....206
Figure 7.1 Human limbal stem cells expanded on CF1-MEF cell layers....211
Figure 7.3.1 PERT assay results on human stem and mouse feeder cell co-cultivation..213
Figure 7.3.2 MLV gag PCR assay results on human stem and mouse feeder cell co-cultivations....214
Figure 7.4.1 Immunofluorescence staining of anti- MLV GAG in human stem cells grown on CF1 feeder layers....216
Figure 7.4.2 Immunofluorescence staining of anti- MLV Gag in human stem cells infected with XMRV....218
Figure 7.5 MLV gag and env relative quantification in co-cultivated stem cells....220
Figure 7.6.1 Splinkerette PCR on genomic extracts from co-cultivated human stem and mouse feeder cells....223
Figure 7.6.2.1 Integration sites from the co-cultivation of stem cells....229
Figure 7.6.2.2. MLV viral sequences integrated into human genomic DNA....232
Figure 7.6.3 Verification of integration sites from genomic DNA of human stem cell co-cultivations....235
Figure 7.7 PERT assay on supernatant from mus dunni cells transfected with stem cell supernatant from co-cultivations....239
Figure 7.8 Phylogenetic tree of MLV isolates in distal mouse ch8 (mm10)....243
Figure Appendix 2.1. Optimisation of the PERT assay: standard curve and background noise strategies....280
Figure Appendix 2.2: PERT assay quantification curve reproducibility....283

LIST OF TABLES

Table 2.1 Cell Lines Used In The Thesis.....62
Table 3.2.1 The Growth Curve Characteristics Of The Cell Lines Used In The Induction Experiments.....90
Table 4.2.6 Syncytia Formation Assays Via Co-Cultivation.....138
Table 5.1 Samples Submitted For Rna-Seq Miseq Ngs 160
Table 5.3.1.1 Mapping To Reference Sequence Results For Cf1 Derived Samples 164
Table 5.3.1.2 De Novo Assembly Results From Cf1-Derived Samples In 169
Rna-Seq 169
Table 5.4 Variant Base Pair Determination In The Cf1-Mlv Consensus.....174
Table 6.3 Datasets Used For The Comparison Of Integration Sites.....199
Table 7.6.1 Output And Mapping To Reference Reads Obtained From Integration Pcr Amplicons 224
Table 7.8 Mouse Chromosome Integration Sequences In Ngs Samples.....242
Appendix Table 1. List Of Primers And Oligos Used In This Thesis 273
Table Appendix 2.1: Commercial Polymerase Endogenous Rt-Like Activity 282
Appendix Table 3 Studies Evaluating 5-Idu For The Induction Of Mlv.....284
Appendix Table 4 Review Of Studies Utilising 5-Azac For The Induction Of Mlv.....288
Appendix Table 9 305
Viral Ltr Sequences Used In Further Mapping Of Integration Sites Of Mlvs Via Ngs.....305

Chapter 1
Introduction

1.1 General introduction

Retroviruses comprise a variety of enveloped RNA viruses that are unique amongst viruses in their replication pattern, which is accomplished via reverse transcription of their viral RNA and ultimate integration as DNA forms into the host genome (Coffin, Hughes and Varmus, 1997; Goff, 2007; Kurth & Bannert, 2010b). Due to this unique life cycle which is instrumental in disease chronicity, they have attracted significant attention during the past 50 years, mainly due to their involvement in the AIDS epidemic that has affected millions of patients worldwide and has driven an unprecedented research effort for their understanding and management (Sharp and Hahn, 2011). Members of these viruses include HIV-1 and HIV-2, HTLV-1 and HTLV-2, murine leukaemia viruses, endogenous human and animal retroviruses and other elements.

Research into retroviruses expanded significantly and continues to expand further over other significant biomedical applications, such as the discovery of oncogenes or the regulation of eukaryotic gene expression; they have also proven significant tools in gene therapy applications as gene vectors (Cosset et al, 1995 ; Kurth and Bannert, 2010b). During the last decade, research interest has sparked over their raison-d- être and the interplay between beneficial or deleterious outcomes stemming from the endogenous retroviral presence and expression; they are described as ancient proviral forms with vast presence in most species but less described functions (Griffiths et al, 2001). One major example of their proven beneficial contribution for hosts however, is the co-opting of a retroviral envelope sequence for the formation of the protein called syncytin, a step that is ultimately responsible for the creation of the syncytiotrophoblast; the latter being absolutely necessary for the implantation of pregnancy in all mammals (Kurth & Bannert,. 2010b , Dupressoir et al, 2011).

Surveillance over the emergence of novel pathogenic retroviruses for humans is constant, while the role of Transfusion Microbiology is considered paramount in ensuring a rapid and co-ordinated response over a potential novel retroviral exposure threat (Urisman et al, 2006; Switzer et al, 2010, 2011; Simmons et al, 2011, Robinson

et al, 2011b). To this extent, the emergence of a murine leukaemia virus recombinant in 2006 reported to infect human blood and tissues, namely XMRV (Urisman et al, 2006), resulted in a swift response from the Blood Service in the UK, and namely NHSBT alongside other public health authorities worldwide for the exact delineation of the associated transmission potential (Tuke et al, 2011; Robinson et al, 2011, Kakisi et al, 2013). The subsequent investigation and risk assessment culminated in assuring funding for the research described in the current thesis.

1.2 Retroviruses

Retroviruses are enveloped RNA viruses that are characteristically diploid and comprise of seven main genera (see Section 1.3: classification). They act counterintuitively to the major biology dogma, transcribing their RNA into DNA and additionally integrate into the chromosomes of infected cells. They all have the basic gene structure of *5'LTR-gag-pro-pol-env-3'LTR*, coding for structural proteins and enzymes flanked by long terminal repeats or LTRs (Coffin, Hughes and Varmus, 1997; Kurth and Bannert, 2010b). However, most retroviruses will encode for additional proteins that have a role at increased pathogenesis. These are called complex retroviruses. The genomic arrangement of simple retroviruses is explained in Figure 1.1.

The structural genes are firstly, *gag* (standing for *g*roup-specific *a*ntigen) which is one of the largest genes and codes for three proteins used in the formation of the viral particle. The gene is transcribed unspliced, as a polyprotein of 80-44 kDa and is further cleaved by the viral protease into the matrix (MA) protein, the capsid (CA) and nucleocapsid (NC) proteins; the matrix protein is 10-19 KDa large and remains bound to the viral lipid bilayer. It participates in late retroviral entry and the process of uncoating as well as in nuclear import and the formation of the pre-integration complex (Hearps and Jans, 2007). The capsid protein (CA) is 20-30KDa large and during maturation, encloses the viral enzymes and the viral genome inside the viral particle core (Provitera, 2001). The nucleocapsid protein (NC) is involved in genomic packaging and tRNA binding during assembly (Carteau et al, 1999). The proteolytic

cleavage of several Gag precursor proteins also produces additional spacer peptides of different sizes in many retroviruses that have a role in early or late stages of the replication process (Accola et al, 1998; Krausslich et al, 1995, Kurth and Bannert, 2010b).

In the *orthoretrovirinae*, Gag, Pro and Pol proteins are all translated from one full length mRNA from which Gag, Gag-Pro and Gag-Pro-Pol polyproteins are synthesised. A successful ratio of translation for all the above proteins (Gag is the most abundantly synthesised, 1:4 is Pro-protease and 1:50 is Pol-polymerase) is achieved by mRNA frame shifting with the formation of hairpin pseudo-knots (Jacks and Varmus, 1985; Marcheschi et al, 2007) and leaky stops. The protease coding sequence immediately follows the Gag coding reading frame. This protease is a 14KDa protein with aspartyl protease activity (Katoh et al, 1987) and has two subunits. Its activation is not fully understood but may involve either pH change during viral release, a high number of concentration molecules needed to achieve dimerisation or conformational changes that will ultimately abrogate inhibitory domains (Goff, 2007).

The *pol* gene codes for the enzyme reverse transcriptase (RT) and the integrase (IN) proteins. Mature polymerase enzymes are cleaved from the Gag-Pro-Pol polyprotein by proteolytic cleavage by the protease. The integrase enzyme is separated by the reverse transcriptase domain using its own protease (Kurth and Bannert, 2010b).

Reverse transcriptase (for the historical discovery see Coffin and Fan, 2016) is the main enzyme responsible for the unique life cycle of retroviruses and is a heterodimer, with an N-terminal polymerase and a C-terminal RNAse H function (Moelling and Broecker, 2015). The active site of the polymerase often contains a conserved YXDD motif. The reverse transcriptase enzyme has low processivity and low fidelity with at least 10^{-3} to 10^{-4} errors per base. It has no proofreading activity. During reverse transcription, the protein can shift from one to the other RNA strand of the viral genome, creating new mosaic genomes via copy-choice recombination. RNase H ensures the degradation of the RNA template some 17-18 bases behind the growing 3-end of the viral cDNA (Gopalakrishnan et al, 1992). A further enzyme, dUTPase was

identified in beta and non-primate lentiviruses (Baldo and McClure ,1999; Hizi and Herzig, 2015) and prevents detrimental uracil base incorporation during the reverse transcription of DNA.

The integrase (IN) protein is a significant component of the pre-integration complex. Its main function is to insert viral DNA into chromosomal DNA through two main actions: the 3' processing of the viral DNA to expose the conserved CA dinucleotides at both 3'ends and the strand transfer reaction, during which processed 3' ends of the viral DNA are ligated to the host chromosomal DNA (Hare et al, 2010).

The *env* gene that codes for the two main envelope proteins, the N-terminal surface (SU) and the C-terminal transmembrane (TM) proteins, is translated from a spliced transcript. The envelope glycoprotein mediates viral attachment and fusion entry into target cells by binding onto receptors of the cell membrane (Goff, 2007; Young 2001). It is the determinant of viral tropism and the target of neutralising antibodies (Burton et al, 2004). TM and SU units remain non-covalently associated and both form trimers that contain the receptor binding region (RBD) in SU, and the 20 amino acid long hydrophobic fusion peptide in TM, that are both indispensable for interaction with the receptor and membrane fusion, respectively (Kurth and Bannert, 2010b; Knopper et al, 2009).

Figure 1.1 Genomic organisation of Murine Leukaemia viruses

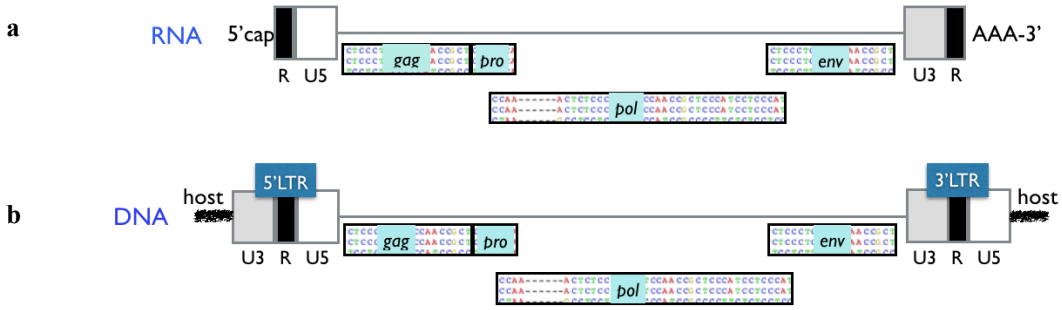


Figure 1.1 Genomic organisation of Murine Leukaemia viruses

The figure shows the basic linear gene arrangement in simple retroviruses including murine leukaemia viruses (MLVs) that are typically 8-9 Kbp long. The RNA form of the virus a) differs from the proviral DNA form (b), mainly in the LTR regions. In the proviral form, these are replicated and consist of two full repeats of the U3-R and U5 regions. The insertion is also flanked at the start of the 5' and the end of the 3' termini by a small duplication of the host DNA, that is 4-5 bp long (not shown) and a 2bp deletion of the viral DNA. Complex retroviruses include additional genes such as *vif*, *tat* or *rev* that contribute to various points of the viral life cycle or aid in pathogenesis.

The retroviral genes are encased between two long-term terminal repeat regions (LTRs) in the 5' and 3' termini in both the RNA and the DNA forms of the virus. The size of these LTRs vary amongst retroviral genera. The LTRs are segmented into three regions, the U3 (Unique 3'), R (Repeated) and U5 (Unique 5') regions. The inverted repeats at the U3 region are responsible for integration as they host attachment sites (*att*) for integrase and aid in 3'-end processing of the proviral substrate (Chow et al, 1992) . The U3 serves as a promoter for cellular polymerase II and once the virus is integrated is the dominant promoter and is responsible for transcriptional interference (Cullen et al, 1984). It also includes several positive and negative regulatory function elements, that vary across genera. These include the Tax protein of HTLV-1, the Tat protein of HIV-1 and the Was proteins of SFV-1 (Cullen, 1992; Kurth and Bannert, 2010b).

The U5 region has been associated with transcriptional enhancing factors as well as viral gene expression, via the action of cis and trans-regulatory elements. At the end of the U5 sequence lies the 18 base-long primer binding site (PBS) that is needed to hybridise with the initiation tRNA during reverse transcription. For endogenous retroviruses, the amino acid specificity of the tRNA primer was originally used to group the viruses into families.

The 3'-end LTR is responsible for transcription termination and for polyadenylation (poly-A') tailing of the nascent RNA forms of the virus. It contains the Polypurine tract (PPT) that is involved in the cleavage of the nascent DNA-RNA hybrid (Huber and Richardson, 1990) prior the start of the plus-strand synthesis. The Repeats region (R) is present at both ends of the termini of the viral transcripts and includes the polyadenylation signal region. The adenine tail of 200-300 bases is added at the end of mature retroviral transcripts to stabilise them and aid in transport, splicing and translation. It is unclear how the 3'-end R region is able to dominate the polyadenylation procedure, possibly through proximity with the terminal cap structure and the downstream splice donor site (Ashe et al, 1997).

1.3 Retroviral virion morphology and classification

The first direct observation of retroviral particles was possible in 1958 via negative stain electron microscopy (Benhard, 1958). The initial observation was that the different genera of retroviruses differ in the appearance of the retroviral virion, the virion assembly site and budding characteristics, and this led to another initial classification of the retroviruses according to their virion morphology.

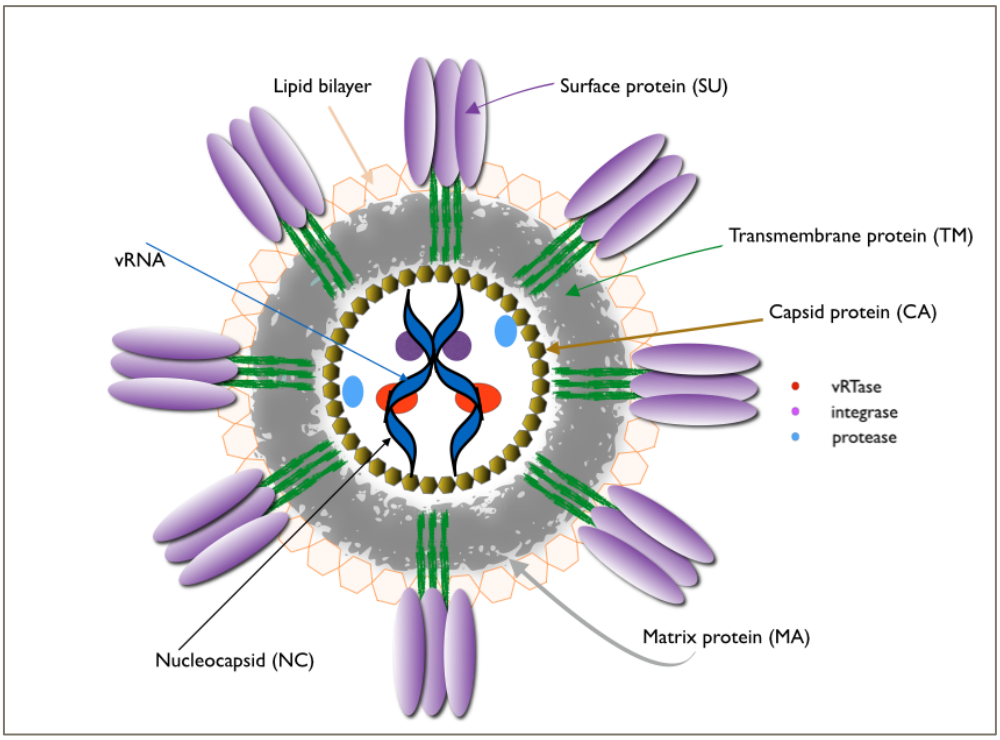
Early EM images showed the clear spike proteins on the surface of the virion. A trimeric structure for the envelope protein subunits was proposed after crystallographic studies (Blacklow et al, 1995; Caffrey et al,1998) and this was proved for human foamy virus particles (Wilk et al, 2000), HIV (Zanetti et al, 2006) and MLV (Förster et al, 2005).

Immediately after assembly, the immature retroviral virion is organised loosely with a spherical morphology of the Gag precursor polyproteins in radial orientation, the matrix protein anchors in the lipid bilayer and the RNA organises in contact with the nucleocapsid (Briggs et al, 2004). The proteolytic processing of the Gag precursors results in the maturation of the particle (Figure 1.3) with the formation of the capsid protein. The two copies of genomic RNA are interconnected at the 5' site (Bender et al, 1978). It is suggested that dimerisation occurs through a multitude of steps, that start with an initial 'kissing complex' that is converted into a stable duplex found in mature virions (D'Souza and Summers, 2005).

The diameter for retroviral virions varies between 80–120 nm in thin-section electron microscopy. The density of retroviruses is approximately 1.16 g/ml, corresponding to about 35% w/w sucrose (Kurth and Bannert, 2010b).

Figure 1.3 Model of the mature retroviral virion structure

A.



B.

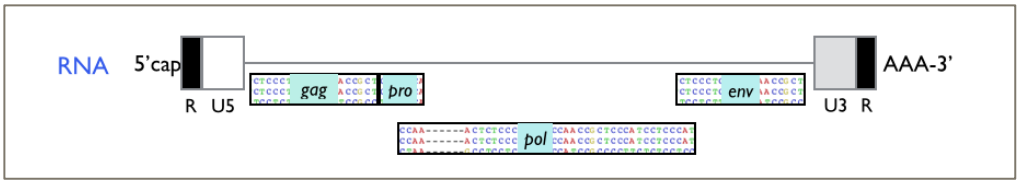


Figure 1.3 Model of the mature retroviral virion structure

A. **Schematic showing a section of a retroviral particle.** The retroviral envelope is formed by the lipid bilayer that originates from the host cell membrane onto which the trimer envelope proteins, surface (SU) and transmembrane (TM) are embedded. Internal structural proteins originating from the proteolytic cleavage of the Gag-Pro-Pol precursor proteins are the matrix (MA), capsid (CA) and nucleocapsid (NC) proteins. In the centre of the core is the complex of NC protein and RNA. Associated with this complex are smaller numbers of reverse transcriptase RT and integrase IN molecules. The functional proteins are also derived from the *Gag-Pro-Pol* precursors through proteolytic cleavage via the action of the viral protease (PR).The protease molecules are shown as two blue dots, because as a mature protein they are a dimer. (author's own-adapted from Retroviruses. Coffin JM, Hughes SH, Varmus HE Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997).

B. **The genomic organisation of a prototype murine leukaemia virus or MLV**.

All retroviruses are grouped under the family Retroviridae. The official taxonomy of retroviruses is based principally on phylogenetic traits determined by analysis of the *pol* gene. This phylogeny has been found to correlate best with the rest of the genetic and structural features: the structure of the virion, the size and number of the virion proteins, the host primer tRNA and/or the presence of additional genes. The seven genera recognised in the family of Retroviridae are mammalian C-type viruses (prototype murine leukaemia virus- MLV), avian C-type viruses (prototype Rous sarcoma virus-RSV), B-type viruses (prototype mouse mammary tumour virus - MMTV), D-type viruses (prototype Mason-Pfizer monkey-type virus M-PMV), lentiviruses (prototype human immunodeficiency virus-1, HIV-1), viruses of the HTLV/BLV group (prototype human T-lymphotropic virus-1 ,HTLV-1), and spumaviruses (prototype human foamy virus, HFV) (Coffin, Hughes and Varmus 1997; Kurth and Bannert, 2010b).

Of the seven genera, five (5) (in detail, avian sarcoma/ leukosis virus, mammalian C-type viruses, B and D type viruses) are recognised as simple retroviruses, characterised by the presence of the common virion proteins encoded by *gag, pro, pol*, and *env* plus at most one extra coding region. A basic characteristic of the simple retroviruses is that they make only one spliced mRNA and that is for envelope. Complex retroviruses include lentiviruses, spumaviruses (foamy viruses), and viruses of the HTLV-BLV group. These contain post-transcriptionally variable products of retroviral accessory genes, which were discovered first in HTLV (Seiki et al. 1983) and regulate the transcription of viral DNA and the processing of RNA. All accessory genes are expressed from singly or multiply spliced RNAs that mature fully or incompletely (reviewed in Cullen, 1992) into their transcripts before export into the cytoplasm. In the simplest cases, the genes are located downstream from *env*, sometimes extending into the U3 region of the LTR.

A further classification concerns the additional existence of oncogenes that are found in Avian retroviruses, Type-C retroviruses and HTLV (oncogenic retroviruses) (Nevins et al, 2007) . These are further classified in the acute-transforming retroviruses that cause acute leukaemia and sarcoma cancers (ASLV, type C) using mainly an

incorporated viral proto-oncogene (v-onc) that transforms cells via growth-enhancing polyclonality leading to cancer rapidly (2-3 weeks). At least 35 viral oncogenes have been described. These acute-transforming viruses are usually defective and not replication-competent. (Rosenberg and Jolicoeur, 1997) Slow-transforming onco-retroviruses (HTLV viruses) on the other hand induce leukaemia over prolonged latent periods. The transforming ability of HTLV-1 for example, has been attributed to the virally encoded oncoprotein Tax that promotes transcription and cell cycle progression. Tax may also lead to chromosomal instability (Marriott et al, 2005), while it is also believed that integration-induced transformation may also play an active role.

The class of retrovirus also plays a significant role in determining the viral infectivity according to the host cells maturation cycle (Katz et al, 2005). Gammaretroviruses such as murine leukaemia virus (MLV) require mitosis for productive infection and are blocked from integrating into non-dividing cells (Roe et al, 1993). These infectivities are not absolute and are usually equally affected by the overall efficiency at infecting non-diving cells; which is equally poor for gammaretroviruses. In contrast, lentiviruses have the highest ability in transfusing both dividing and non-dividing cells (Lewis et al, 1992). Similarly, novel research suggests there might be many mechanisms that account for this viral property (Yamashita and Emerman, 2006).

1.4 Murine Leukaemia Viruses (MLVs)

1.4.1 MLV structure and replication

Murine Leukaemia viruses (MLVs) are endogenous retroviruses of mice. Their virions are 80-100 nm in diameter and their outer lipid envelope contains their viral glycoproteins. (Moloney 1960; Sutcliffe et al,1980). The virions have a condensed central core with surface spikes. For replication- competent MLVs the genome size of the viral RNA ranges approximately from 8 to 8.5 Kbp (Figure 1.1).

Mature virus particles of Murine Leukaemia Viruses are recognised by their electron-dense C-type cores, which consist of polyhedral lattices of the CA, beneath which reside the dimeric RNA molecules and RT/IN enzymes, encapsidated by NC (D'Souza et al, 2004). (Figure 1.3). The viral cycle of infection for MLV viruses is shown in Figure 1.4.1 (A-L) (adapted from Goff, 2007). The virus first attaches to the cell receptors that are compatible with its particular envelope protein (A), the fusion of the membrane at the adsorption site then causes stereotactic changes that in turn, facilitate the entry of the viral particle contents into the cell cytoplasm (B).

The viral RNAs and enzymes post-cell entry are released and RNAs are reverse transcribed into double strand DNA (C). The viral DNA is transported into the cell nucleus (D), where it integrates with the aid of viral integrase into chromosomal DNA (E). It can now be transcribed (F) into genomic or viral RNAs (F) that exit the nucleus (I) or into spliced messenger *env* RNA (G). The messenger RNA exits the nucleus and is translated into Gag-Pro-Pol precursor viral proteins in the ribosomal units of the cytoplasm (H). The new viral particle consists initially of genomic RNA and the Gag-Pro-Pol proteins arranged in the viral core in a radial fashion (immature particle), with the envelope protein subunits SU and TM arranged as trimers linked by disulphide bonds at the outer layer. This assembly (J) precedes the release of the virion from the cellular membrane through budding and acquisition of a lipid bilayer (K) while proteolytic maturation, via the action of the viral protease, gives rise to the capsid (CA), matrix (MA) and nucleocapsid (NC) proteins and the viral enzymes reverse transcriptase (RT) and integrase (IN), that all together constitute a full mature virion (L).

Figure 1.4.1 Replication cycle of MLVs

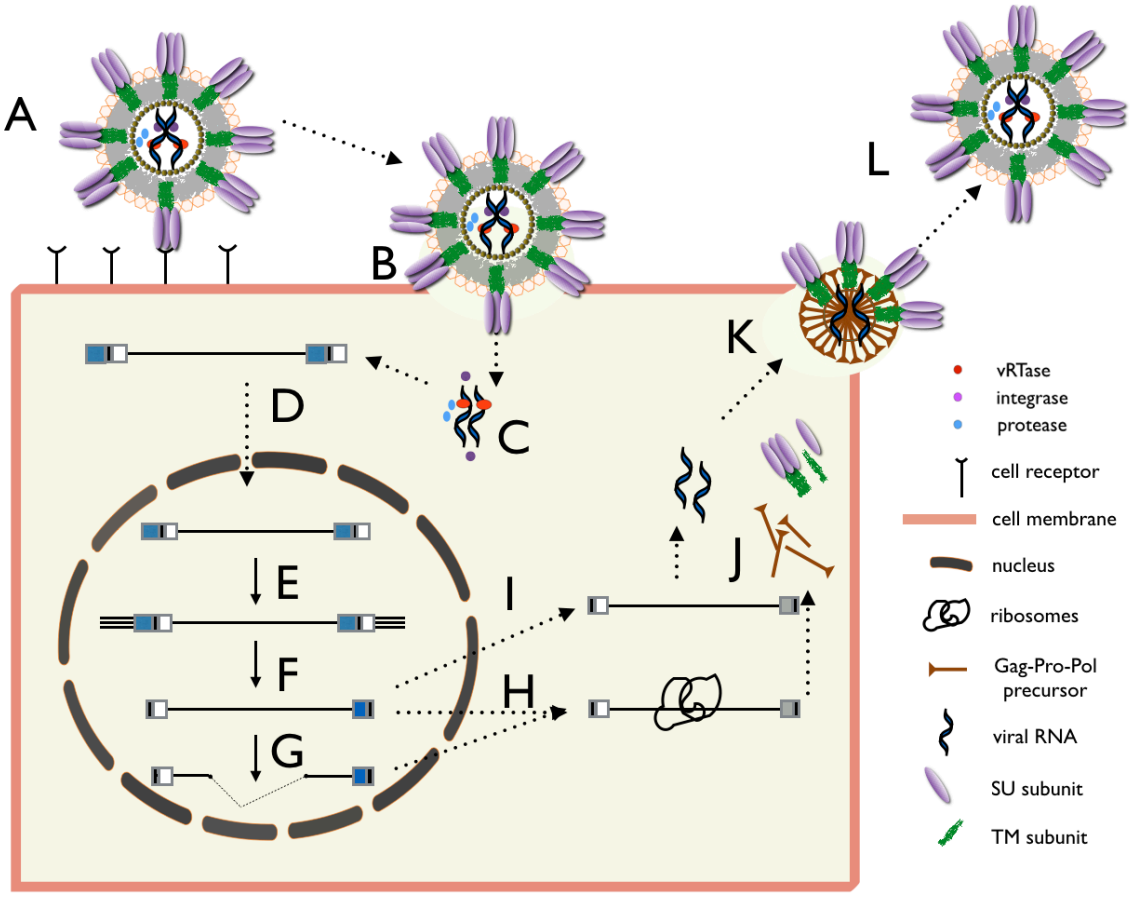


Figure 1.4.1 Replication cycle of MLVs.

The steps of the processes involved in MLV infection are shown as:

A. Cell receptor recognition and binding
B. Host cell entry via membrane fusion
C. Reverse transcription
D. Nuclear translocation of viral DNA
E. Integration in chromosomal DNA
F. Proviral transcription
G. Spliced mRNA transcription
H. Protein translation
I. gRNA transcription
J. Assembly of genomic RNA and virion protein precursors
K. Release by budding
L. Proteolytic cleavage and maturation of full virion

1.4.2. Classification of MLVs

1.4.2.1 Endogenous and exogenous MLVs

Endogenous are the proviruses that are stably inherent to the cell genome and are transmitted vertically from mother to offspring. They are considered remnants of ancient species infection and are in their majority inactive (for a review see Weiss, 2006). For MLVs, they can be replication-competent or defective, while mice bear a significant number of either types. This is in contrast to other mammalian species, including humans, where endogenous retroviruses (ERVs) appear to be mainly defective and thus replication incompetent (for a review and exceptions see review by Stoye, 2012).

Exogenous retroviruses are the traditional focus of clinical virology as many endogenous viruses do not readily re-infect their own cells; the exogenous (from Greek exo=outside) viruses that aim at infecting new hosts are replication-competent, infectious and transmissible horizontally, mainly through contact with blood and bodily fluids. It is sought that endogenous infections are the result of an ancient exogenous infection that 'endogenised' over a multitude of years inside a host genome and particularly the host germinal cells. An example of such an endogenous retrovirus establishing germ-line transmission during recent times, is the Koala Endogenous Retrovirus (Ko-ERV) (Stoye, 2006). Indeed, host switching is also widely described for most of the known endogenous MLVs (Martin et al, 2003).

1.4.2.2 Classification of MLVs according to receptor usage

The MLVs can be further classified into several subgroups by distinction of their *env* gene, which is responsible for the receptors that the viruses ultimately recognise for host cell entry and viral adsorption. This classification is linked to the initial observation that the SU (surface) of MLV envelope in its soluble form is responsible for receptor binding (De Larco and Todaro, 1976). Surface (SU) envelope protein is encoded in the amino-terminal region of the envelope open reading frame (ORF) and is translated from spliced subgenomic *env* mRNA. It is

now known that the SU protein is involved in two structure-function relationships: a) receptor recognition via the receptor binding domain (RBD) and b) fusion signalling prior to fusion pore formation (mediated by TM) (Davey et al, 1997; Kurth and Bannert, 2010b).

Based on the differences in SU and subsequent receptor use in mouse cells there exist three types of naturally- occurring MLVs: ecotropic, polytropic and xenotropic (Stoye& Coffin, 1987; Kozak et al, 2011)). The ecotropic viruses infect only murine and rat cells and use the cationic amino-acid transporter receptor (mCAT-1) (Davey et al, 1997). The xenotropic viruses infect non-rodent hosts only, while polytropic viruses infect both mouse and non-rodent hosts; both use the XPR-1 receptor (Kozak et al, 2011). One other class of MLVs, the amphotropic MLVs are isolated from feral mice and although do not produce disease in their host cells, are able to infect other species host cells (using the Ram-1 receptor) (Hartley and Rowe, 1976; Rasheed et al, 1976). Many human cells are susceptible but their exact host range has not been determined.

The coding regions of the surface (SU) envelope protein, that are most highly responsible for receptor recognition and are variable amongst MLV isolates are shown in Figure 1.4.2. Receptor specificity is determined mainly by the receptor binding domain (RBD) which includes three regions (Kozak et al, 2011): variable region A (VRA), variable region B (VRB) and variable region C (VRC) (Figure 1.4.2A). The RBD is followed by the proline rich domain (PRD) that separates the amino-terminus variable region from the highly conserved C-terminal domain, near the start of the transmembrane protein (TM) coding region.

Viruses that belong to any of the groups are able to interfere with each other, especially in their endogenous forms. The ecotropic viruses' envelope genes differ the most, from a phylogenetic point of view, from those of the other two subgroups and are also quite variable relative to one another (Koch et al, 1983). Compared to the rest of the isolates, the ecotropic MLVs have a large insertion in

VRA, a deletion in VRB and numerous variations in the proline rich region (Kozak et al, 2011).

Deletion or insertional mutagenesis of the VRA/VRB region is able to disrupt receptor recognition completely (Heard and Danos, 1991, Gray and Roth, 1993). The receptor binding domain, comprising of the 245 amino-terminal amino acids determines the specificity for the ecotropic receptor, mCAT-1. Indeed, replacement of the VRA and VRB in an ecotropic MLV virus with the respective VRA and VRB regions of an amphotropic virus (4070A) alters viral specificity for the ecotropic receptor and confers the specificity of the amphotropic recombinant (Morgan et al, 1993).

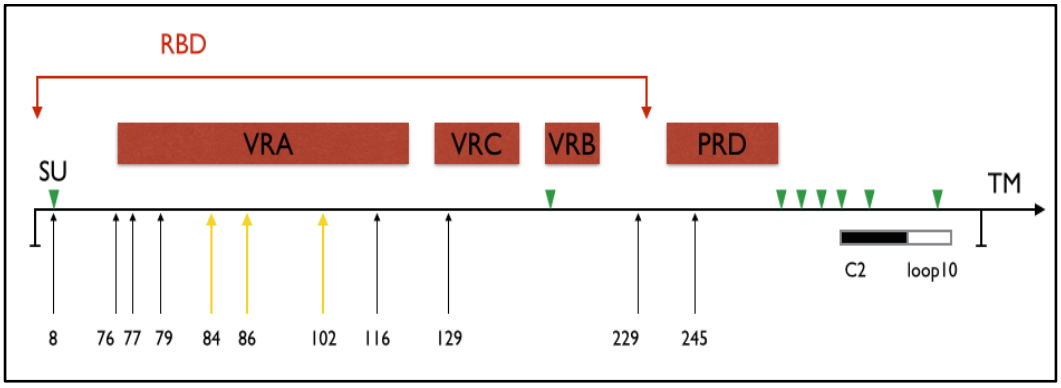
Deletion of the proline-rich domain (PRD) however, does not abolish the ability of the soluble protein fragment to induce ecotropic virus interference, to the same level as that of the intact ecotropic envelope glycoprotein (Battini et al, 1994). The PRD is found to be involved in post-binding conformational changes, mediating fusion (Lavillette, 1998) and is considered to influence syncytia formation (Ferrarone et al, 2012).

From a structural view, the mCAT-1 recognition sequences lie within the VRA of the RBD. Three (3) particular residues have been identified in Friend-e-MLV (Kozak, 2011) as critical for viral entry: S84, D86 and W102 (Figure 1.4.2B). This is considered to occur due to their central position in the 'binding pocket' of the secondary structure of the surface envelope protein (Figure 1.4.2B author's own; in yellow). Additionally, several post-binding entry-defining residues have also been described, such as the H8 residue and the C-terminus which are involved in fusion (Kozak, 2011).

For xenotropic and polytropic viruses, an amino-terminal region replacement (VRA) is sufficient to determine receptor usage specificity (Battini et al, 1994), however the critical aminoacids for XPR-1 binding have not yet been identified. Xenotropic and polytropic isolates are also very variable in the regions containing the RBD.

Figure 1.4.2 Structures of the Surface Envelope MLV (SU Env-MLV) protein that determine virus tropism

A.



B.

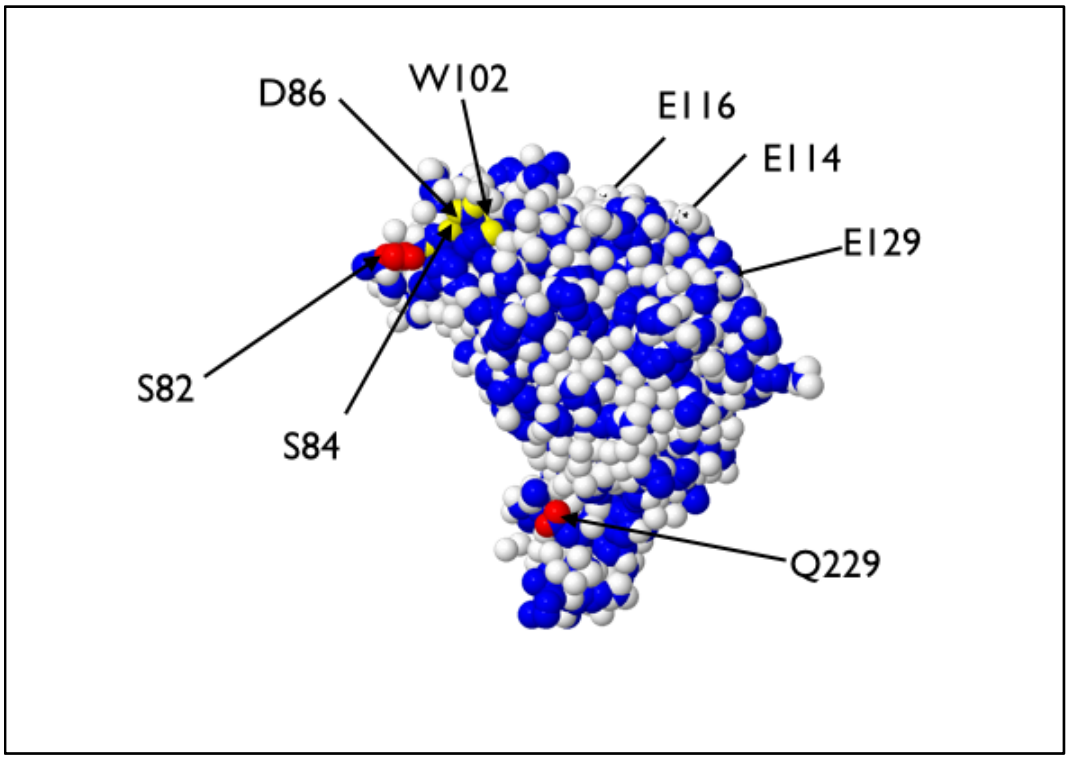


Figure 1.4.2 Structures of the Surface Envelope MLV (SU Env-MLV) protein that determine virus tropism

A. The figure illustrates linearly the anatomy of the surface (SU) protein of a prototype ecotropic virus (Friend MLV), comprising of the RBD, that contains the three variable regions VRA, VRB and VRC followed by the Proline-rich domain (PRD). In straight arrows are marked the positions of the viral entry determinant amino acids (AAs). In yellow are the three "binding-pocket" central amino acids that are primarily important for viral entry. Green arrows mark the N-linked glycosylation sites in SU. The C terminal segments C2 and loop10 have also been implicated in viral entry (adapted from Kozak, 2011).

B. The schematic representation of the 3-D structure of the Friend MLV SU protein (Protein Database ID: 1AOL). In blue are protein carbon residues, while custom yellow shows the location of the binding pocket amino acids and red marks additional residues involved in receptor specificity as in Kozak et al,2011, as also shown in figure 1.4.2A.

1.4.3 Recombination in MLVs

A characteristic of MLV and all retroviral replication is the ability for genetic recombination, a genetic rearrangement capacity between the two genomic RNA copies that has been known for several years. It is an issue of additional clinical concern in HIV replication, as recombination in HIV viruses has impacted greatly the dynamics of viral populations in patients, with more than 10% of circulating viruses being recombinants of circulating viral isolates (a phenomenon attenuated by superinfection with different viral strains leading to the formation of heterozygous virions) (Sharp et al, 1999; Blackard et al, 2002).

As retroviruses are diploid, having two RNA molecules in each virion, it was originally hypothesised that recombination occurs via exchange of genetic information between the two co-packaged RNA molecules, due to their close proximity. Hu& Temin 1990 demonstrated that retroviral recombination only occurs in heterozygous virions, i.e. virions containing two genetically distinct genomic RNAs in the same viral particle. Using a single-infection cycle experimental system, they showed that recombination takes place during reverse transcription. As the reverse transcriptase synthesises the first (-) cDNA strand (Luo and Taylor, 1990; Hwang et al, 2001), the enzyme is able to switch templates from one copy of the RNA to the other, a process known as copy choice (Coffin, 1979). The final cDNA molecule (Figure 1.6.B) consists of segments from both the genomic RNAs inside the heterozygous virion (parental RNAs), i.e. is a recombinant cDNA.

This recombination affects all retroviral genes and concerns mainly homologous gene regions, i.e. *gag* to *gag* etc. Indeed, when sequences are not homologous the rate of recombination falls by 100-1000 fold (Zhang and Temin, 1993). This is explained by the close affinity of homologous regions that shifts the donor RNA templates structurally closer to sites of identity, facilitating template switching of the RT enzyme. Indeed, in cases where recombination is strictly unwanted i.e. the use of gene therapy with MLV- based vectors, the presence of overlapping nucleotide sequences between the vector and the packaging constructs leads to a higher incidence of

recombinant formation of replication- competent retroviruses (Shin Yu et al, 2003). The cleaving of the copied RNA strand once it has been copied depends on the activity of a ribonuclease (RNAse H) carried by the reverse transcriptase enzyme (Peliska and Benkovic, 1992).

Overall determinants of template switching apart from the close homology in nucleotides, are: a) the 5-termini and b) any strong hairpin cis-acting structures (Topping et al, 1998) in donor templates (preferred sites of recombination) and c) junction sites. Triggers for the template switch of the RT enzyme may be: discontinuities in the parental RNA templates (causing escape-lane rescue synthesis) (Coffin, 1979), difficulties in incorporating a particular nucleotide (either in the negative strand cDNA or the positive strand) that lead to enzyme pausing (Lanciault and Champoux, 2006) and any close affinity between acceptor RNA and the catalytic site of the RT enzyme (Kim et al, 1997). Indeed, when active competition of the acceptor and donor RNA with the enzyme occurs, the RT molecule is forced to switch templates more frequently (via active displacement). Last, the RNA-binding nucleocapsid protein (NC) is also known to enhance strand transfer by modulating strongly structured regions of the template- thus, creating hotspots for switches (Negroni and Buc, 2000).

The frequency of recombination is determined by the frequency of the cross-over events during template switching. For HIV-1, the recombination rate has been measured at 2.4×10^{-4}/ bp per replication cycle or higher (Rhodes et al, 2003) and for MLV at 1.09×10^{-5} /bp per replication cycle (Parthasarathi et al, 1995). This difference is nowadays considered less prominent, as for homologous recombination, MLV and HIV have similar recombination rates (Varela- Echavarria et al, 1990).

One significant difference between MLVs and HIV lies in the ability of MLVs to recombine between exogenous and endogenous proviruses that are found abundant in mice, thus allowing the emergence of a much greater phenotypic and genotypic variety. This was first observed in mouse tumours, where recombination with endogenous viruses leads to the birth of more pathogenic strains (Stoye et al, 1991). In

NSF/N mice, 3 endogenous proviruses are found to be major participants in the recombination with Mo-MLV (Moloney MLV). It has been shown that exogenous infection has the ability to mobilise intact endogenous virions (Evans et al, 2009), however complementation via the exchange or 'lending' of envelope sequences seems to be the most frequent occurrence. This process of incorporation of homologous or heterologous envelope sequences by viruses is otherwise known as pseudo typing (Zavada, 1982).

The recombination of endogenous MLV viruses is clearly demonstrated in the emergence of Mink-cell forming (MCF) viruses in AKR mice (Vogt et al, 1979; Elder et al, 1977). AKR mice are naturally infected with a high leukaemogenic strain of an ecotropic endogenous MLV, the AKV virus. AKV is produced in AKR mice and is present in high concentrations. Persistent infection with the virus leads to the stimulation of endogenous polytropic sequences during the pre-leukaemogenic stage, that recombine with AKV forming viruses covered with recombinant *env*, which is polytropic (Stocking and Kozak, 1998). The novel viruses, are highly pathogenic and increase the severity and rate of progression towards leukaemia via the activation of oncogenes. MCF viruses have a distinctive 27 bp deletion in *env* and duplications in the U3 enhancers in the LTR.

The ability of retroviral *env* recombination among endogenous retroviruses leads to several interesting natural phenomena, such as 'the species-jump' noted in the case of an Asian rodent retrovirus that became pathogenic in gibbons and koalas (Kurth & Bannert, 2010a). It also lies behind the process leading to the emergence of xenotropic isolates, with ability or not to infect human cells such as XMRV (see next section) or BXV-1, an endogenous xenotropic virus of Balb/c mice; and this has also been described in other than mice species, as is the case with RD-114, a xenotropic retrovirus that arose as a result of endogenous retrovirus recombination in cats during xenografting conditions (van der Kuyl *et al.*, 1999). More recently, an established ecotropic MLV viral infection in mice was studied in Rag-/- immunocompromised mice, to generate a high amount of recombinant viruses mobilised by a replication-competent endogenous ecotropic MLV (emv-2) (Young et al, 2012).

1.5 The emergence of XMRV

(from the original publication Kakisi et al, 2013)

In 2006, researchers at the Cleveland Clinic and the University of California, San Francisco, (Urisman *et al.*, 2006) found a novel *Gammaretrovirus*, XMRV, in 40% of prostate tumours from 86 men with familial prostate cancer (PCa) that was characterised by a genetic defect in an antiviral resistance gene, *RNAse L* (Bisbal & Salehzada, 2008). This was done using a novel, gene array, the viro-chip (Chen *et al.*, 2011) that operates via the interaction of patients' tissue DNA with a wide range of immobilised viral molecular probes.

The virus had close homology to murine endogenous retroviruses but was a distinctively new entity. The identification of XMRV as *Gammaretrovirus sp.* and its localisation inside human cells surprisingly did not attract much attention or concern, until 2009, when the presence of XMRV was re-examined in prostate cancer (PCa). It was then reported that PCa patients with a normal *RNAse L* status were also 6% positive for XMRV DNA and 23% were also positive for XMRV protein by immunohistochemistry, while XMRV positive tumours were associated with a higher grade of malignancy (Schlaberg *et al.*, 2009).

XMRV was subsequently fully sequenced and integration sites were demonstrated in PCa cells' genomic DNA (Kim *et al.*, 2008), a hallmark of true retroviral infection. XMRV was an exogenous virus, transmitted effectively in human cells *in vitro* (Stieler *et al.*, 2010) using the membrane receptor XPR-1, as a point of entry specific for xenotropic and polytropic murine retroviruses (Kozak, 2010).

Chronic fatigue syndrome (CFS; for a review see Ranjith, 2005) is a debilitating disease causing extreme fatigue and variable neurologic deficits (also known as myalgic encephalitis or ME). Some CFS patients also carry a defective *RNAse L* gene, an observation that initially led investigators to look for a role for XMRV in the aetiology and pathophysiology of CFS. CFS/ME syndrome had also been previously linked to post-viral fatigue, which made the implication of XMRV infection plausible.

In 2009, a research group from the Whittemore Peterson Institute (WPI) for Neuroimmune Disease, led by Judy Mikovits, published a paper in *Science* demonstrating the presence of XMRV in the peripheral blood mononuclear cells (PBMCs) of 67% of patients with CFS and in 4% of healthy study controls. This was seen as a major step forwards in the understanding of CFS, not only bringing an etiological coherence to an orphan disease, but more importantly offering the hope of a therapy in the future.

While disease associations with PCa and CFS were being actively explored, concern was expressed in public health circles about the detection of XMRV infection in the healthy donors used as controls in the Lombardi *Science* publication (Lombardi *et al.*, 2009). An emerging retroviral infection, potentially transmissible by blood transfusion, was a chilling possibility. Although patients with cancer (including PCa) are excluded from blood donation, many early cases are likely to be asymptomatic for many years and thus, the potential of transmission from male blood donors who might be harbouring early undiagnosed disease was raised. All known human retroviruses are so far considered to have emerged from primates. However, mouse and human contact is not an implausible opportunity in many parts of the world whether urban or rural. An emerging retroviral zoonotic infection thus, could very well be an addition to the list of other rodent infections that are known to pass into humans.

With this background in mind, shortly after the Lombardi *et al.*, study, the AABB (American Association of Blood Banks) established an XMRV inter-organisational task force. At the same time, the 'Blood XMRV Scientific Research Working Group' formed under the auspices of the US government with the purpose to formally inform the debate over the possibility of the virus spread in the human population, its pathogenic potential and the risk to the blood supply. In April 2010, several countries with the lead of Australia (www.donateblood.com.au), New Zealand (www.voxy.co.nz) and Canada (www.cbc.ca) decided to ban blood donations from previously affected CFS patients, as a proportionate response to the possibility of a cryptic infection. Concern by the Blood Services was reasonable and legitimate, as the

decision to refuse blood donations is a duty of care justified over concern for blood safety.

Quite soon after the publication of the *Science* paper in 2009, increasing numbers of studies during 2010, consistently failed to replicate its findings across a wide geographical area in the UK (Erlwein *et al.*, 2010, Groom *et al.*, 2010), the Netherlands (van Kuppeveld *et al.*, 2010) and the United States (Switzer *et al.*, 2010); while researchers extended testing to a wide variety of orphan or obscure disease, such as other manifestations of cancer and autism disorders where results repeatedly proved negative. The discussion focused on the divergent methodologies used to culture, extract and amplify viruses from clinical samples. Shortly after the publication, NHS Blood and Transplant (NHSBT) and its advisory committee, the Standing Advisory Committee on Transfusion Transmitted Infection (SACTTI), drew up the first risk assessment from published information for the potential risk to the blood supply in the UK. This assessment was passed to the Joint United Kingdom Blood Transfusion Services and National Institute of Biological Standards and Control Professional Advisory Committee (JPAC).

During the spring of 2010, studies were undertaken by NHSBT to develop a molecular detection method [one step quantitative polymerase chain reaction (PCR)] complemented by sequencing to detect XMRV genomes in whole blood and plasma RNA. A preliminary investigation of the prevalence of XMRV in UK blood donors using this protocol was initiated. A series of 677 whole blood samples were assayed and no infected donors were identified.

Over the following few months, a sensitive real time PCR (RT-PCR), similar to the one currently used for Hepatitis C virus (HCV) RNA detection, was developed for the detection of XMRV virions in plasma. The assay was used to investigate 480 plasma mini pools of 48 samples (representing 23 040 individual donations) and found no evidence of XMRV infection (Robinson et al., 2011). At the same time PCa tissue samples were also tested by the same method. Conflicting data arose with the detection of PCR signals from the DNA extracts of these sections, and the failure to

detect XMRV genome in blood samples taken from the same patients. While SACTTI and JPAC regularly reviewed updates, the NEPNEI (National Expert Panel on New and Emerging Infections) held an XMRV subgroup meeting in March 2010 where it was concluded that, until further evidence could be generated, no further public health action would be required.

Later in the summer of 2010, as attempts were made to comprehend the conflicting findings, PCR 'no template' (water) control samples, as well as samples from empty DNA purification columns and master mix reagents, could spontaneously give positive signals in the highly specific and sensitive assay for MLVs; showing evidence for the first time of contamination by murine sequences in PCR reagents and other components used for routine viral detection (Tuke et al., 2011). By the end of summer 2010, the international controversy still showed no signs of abating but was conversely further fuelled in August, when scientists from the Food and Drug Administration, National Institutes of Health and Harvard Medical School published their detection of MLV polytropic retroviral sequences, closely related to XMRV, in the blood of 86·5% patients with CFS and in 6·8% of healthy individuals (Lo et al., 2010). This at once re-ignited the concern about the risk of XMRV circulating in the healthy population.

In September 2010, an International Workshop held in Bethesda, MD, USA brought international experts together in an attempt to clarify all this divergent data. The number of negative reports that failed to identify XMRV had already begun to dent the credibility of a disease association with CSF/ME and started to question the credibility of an association with PCa. Although a considerable body of association data was aired, the meeting was pretty much bipolar with a division between those who were able to identify XMRV sequences in patient materials and those who could not. The issue of the proficiency of cell culture in detecting the virus was a major bone of contention, as it was controversially being put forward by the WPI researchers as the 'gold standard' for XMRV, over molecular methods.

During proceedings, unpublished data were presented by Dr Judy Mikovits from the WPI Institute, which indicated that around 5% of English blood donors were XMRV-

positive. The provenance of samples used for this investigation was never established. Furthermore, no records were ever found to substantiate the transfer of samples from NHSBT donors to the WPI and subsequent data was published to refute this possibility (Robinson et al., 2010). Furthermore, initial results from testing for XMRV antibody on the widely used Architect platform were also reported in an increasing number of US donor samples and yet presented with failure to identify any donor as seropositive or 'reactive' over a range of XMRV proteins.

Geographical differences in the prevalence of XMRV infection in PCa groups in further American, European and global studies were attributed to discrepancies between detection assays, though the mechanism behind this remained at the time, unclear (for a full review see van Kuppeveld & van der Meer, 2012 and in PCa, see Sfanos et al., 2012). However, in vitro studies showed rather quickly that the virus was susceptible to inactivation by antiviral restriction proteins namely human APOBEC and tetherin (Groom et al., 2010), both actively produced in healthy human cells in response to a retroviral infection. This observation made the probability of a widespread infection in immunocompetent individuals rather unlikely. Traditional antiretroviral medication was also shown to inhibit the virus *in vitro*. At this point, anecdotal reports of CFS/ME patients self-prescribing antiretroviral therapy for XMRV, raised significant worry in medical and ethical circles.

In November 2010, NHSBT issued a deferral for CFS/ME patients, excluding them from blood donation on the basis of donor safety in line with the EC directive, in order to avoid any further compromise over the fragility of their health. NHSBT acknowledged at the time, the lack of data on whether XMRV was a true blood borne viral infection, but decided to keep the issue under review.

1.6 XMRV and the contamination origin

1.6.1 XMRV-related contamination

In December 2010, four research articles and an editorial appeared simultaneously in the journal Retrovirology (Hué et al., 2010; Oakes et al., 2010; Sato et al., 2010; Smith, 2010), presenting evidence supporting that XMRV was indeed a contaminant from PCR components, laboratory material and cell culture. It transpired that the cell line 22rv1, a prodigious secretor of XMRV, had been widely shared by laboratories working on prostatic disease in the late 1990s and may have been present in those laboratories involved in the initial discovery of XMRV.

The scientific community now, turned its attention once more to the Lombardi-CFS article which was being continuously defended by its authors and focused attention on the output of several of the experiments described therein. Questions were raised with respect to the selection of cohorts of patients in the original report, as these were considered not representative and the experiments neither randomised nor blinded (Lloyd et al., 2010).

Many questioned the demonstration of a single peak in the flow cytometry profile of infected blood cells, as this is inherently unlikely in any virally infected cell population, either from tissue culture or blood, as a small percentage of cells will always remain uninfected or at least un-expressing even in the most productive of retroviral infections. Lack of antibody responses in a large number of ‘virus-positive’ subjects (Robinson et al., 2011) and scientific misconduct in the open such as the infidelity of published graphics to their attributed legends led to the question of the overall validity of the association papers.

Indeed, patient samples from the original Lombardi and the Lo et al., studies were re-tested and found to be all negative in individual (Knox et al., 2011, Shin et al., 2011) and multiple laboratories (Simmons et al., 2011) and last, in a blinded randomised fashion (Alter et al., 2012). All the above led to an instinctive retraction of the paper

by Science (Alberts, 2011) and the later retraction of the supporting Lo et al., (2012) paper and Schlamber et al 2009.

In June 2011, Paprotka et al., reported of the probable recombination origin of XMRV from two endogenous murine retroviruses in the 22rv1 PCa cell line which was serially passed through nude mice (Figure 1.6). The authors discovered evidence of a 99% similarity in nucleotide bases between the putative construct of two endogenous murine proviruses, pre-XMRV-1 and pre-XMRV-2 and, XMRV-VP62 as reported in the initial PCa report by Urisman et al., 2006.

1.6.2 Disease dis-association

The 'contamination origin' of XMRV as a human pathogen was strengthened via the accumulation of studies dissociating XMRV from a causal role in CFS/ME. These negative findings, the result of efforts and huge costs put forward from researchers and institutions worldwide, were the first indications that XMRV was not a human pathogen. Precedence for mouse viral sequences accidentally detected in cell lines is documented (Nicklas *et al*., 1993), as similar endogenous retrovirus contamination had created problems to laboratories in the past. The reporting of spontaneous (Aaronson *et al*., 1969) or chemically induced (Stephenson *et al*., 1974) rescue of endogenous type C retroviruses in cell lines has been vast since the years of the boom of cancer research in the 1970s.

Given the degree of specificity of molecular tools available, the finding of the exact sequence match of XMRV isolates between patients and cell lines was solid evidence against an already established infection in man (Hué *et al.*, 2010; Garson *et al.,* 2011; Rusmevichientong *et al.*, 2011). In reality, *in vivo* retroviral infections create important sequence variation between hosts, either due to the selective pressures exerted by the host innate immune system and/or the predictable recombination and mixing of circulating infectious viruses, all providing genetic heterogeneity that was not seen in XMRV phylogeny from 'infected' patient samples.

Perhaps the most telling observation supporting the contamination origin was the further description of identical genome integration sites in a number of prostate tissues bearing the XMRV provirus, (Rusmevichientong *et al*., 2011) purporting to be independent isolates and later on, the identification of identical SNPs (single nucleotide polymorphism). This level of identity can be explained by a contamination event with high confidence.

1.6.3 Mouse DNA and MLV contamination

The prevalence of murine retroviruses in materials that are of interest to biomedical research is astonishing. Since the discovery of the recombination origin of XMRV in nude mice, Cingöz *et al*., 2011 have screened 94 wild and laboratory mouse strains for the presence of its two parental proviruses and found them together in only three laboratory strains, validating the origin of the full virus from inside the laboratory. At present it is documented that 26% of xenograft-derived and 17% of xenograft-free human tumour cell lines (Zhang *et al*., 2011), common inbred mice strains (Baliji *et al*., 2010), commercially available human cell line DNA and a very common strain of fibroblast layer used as a feeder cell line for a variety of human stem cell applications (NIH/3T3) (Mendoza *et al*., 2011) all contain either various polytropic or xenotropic MLV proviruses, XMRV or one or both of the pre-XMRV proviruses. To add to further confusion, common molecular laboratory reagents have now been shown to harbour murine contaminating DNA (Erlwein et al., 2011a,2011b, Wolff & Gerritzen, 2011; Zhang *et al*., 2011).

Normally, but not invariably, viral infection of cell lines produces a change in the morphology of the cells, termed a cytopathic effect (CPE). Gammaretroviruses, however are often non-cytopathic, producing no detectable CPE. Therefore, their presence in tissue culture can remain unnoticed. This can compromise valuable results of cumbersome and costly research. Even the simplest of cell culture related research hypothesis can be compromised by the presence of MLVs (see Lusso *et al.*, 1990, Takeuchi et al., 2008a, b, Weiss, 2010). Promising new treatments such as the ones employing the use of retroviral gene therapy vectors are in need of rigorous monitoring, as any inadvertent mobilisation of endogenous viruses can be detrimental.

Highly sensitive product-enhanced reverse transcriptase assays that measure retroviral contaminant activity are the official screening test for a spectrum of biological products, such as monoclonal antibodies and pharmaceuticals. This and further standard tests established in governmental health and safety offices as well as in pharmacopoeia serve as the barrier to contamination until now. However, any biomedical experimentation and research (cell research for academia for example) that is uncontrolled for such contamination is subject to error, questionable applicability and most of all, questionable translation value.

A rigorous risk assessment has found no risk of MLVs to the safety of live attenuated vaccines (Switzer *et al*., 2011). Screening for possible transmission to animal and lab workers has also proved negative (Brooks *et al*., 2012).

XMRV can now be generally regarded as a salutary modelling exercise in current virology surveillance; monitoring the response speed and the effectiveness of the reflexes of the scientific community in the process of a new virus emergence in the 21st century. Novel techniques and recently developed PCR methods that are widely used for virus detection were compared and scrutinised and an extensive mouse contamination in laboratories was finally made clear to the research community. New techniques and new collaborations were developed. It was unfortunate, however, that scientific conduct in light of a new virus emergence was perhaps less critical that required; and this offers significant lessons in this new era of open communication.

1.7 Human stem cell co-cultivation with mouse cells

1.7.1 The principle of human and mouse cell co-cultivation

Cell therapy for human disease has bloomed in the 1980s after the discovery that murine embryonic fibroblasts (MEF) are able to sustain the growth of human keratinocytes cells in culture. Previously, the growth of human epidermal-like cells *in vitro* was poor. However, the observation was made (for a narrative of the discovery see Green, 2008) that direct contact with a confluent irradiated layer of murine MEF

cells promotes human keratinocyte growth to the desired extent i.e. the production of epidermal sheets that are suitable for transplantation. Hence, this method provided the basis for several types of transplant products such as cultured epidermis for burn victims and limbal stem cell transplants for ocular surface disease. The basic protocol used in direct contact cell culture of stem cells and murine feeder layers (MEFs) is shown in Figure 1.7.1.

The arising cultured-cell products are classified by the regulatory authorities as xenografts and significant concern has been raised, mainly over the risk of remaining mouse cells in the final human cell graft, that can be transplanted to the human host and may harbour the risk of a murine virus transmission. Indeed, the recent emergence of XMRV in serial xenografts of human prostate cancer cells in nude mice *in vivo* has highlighted the actual possibility of not only transmission of murine retroviruses; but also novel genesis by recombination of a pathogenic to human cells, virus isolate. This hypothesis was solidified by the demonstration of recombinant xenotropic virus formation in Rag-/- immunocompromised mice, that generate a high amount of recombinant viruses mobilised by a replication-competent endogenous ecotropic MLV (emv-2) (Young et al, 2012). It is now known that the practice of stem cell co-cultivation has shifted towards xenogeneic free conditions and several other methods have been developed in place of mouse feeder sustainment such as amniotic membrane (Grueterich et al, 2002), explant co-culture , mesenchymal feeders (Omoto et al, 2009), or animal-free serum media supplements.

The principal aim of this work is therefore, to determine the possibility of activation of replication-competent murine leukaemia viruses (MLVs) from the inactive endogenous reservoir present in MEF cells that are used in stem cell co-cultivations pre-clinically.

Figure 1.7.1 Stem cell co-cultivation with mouse feeder layers: the principle

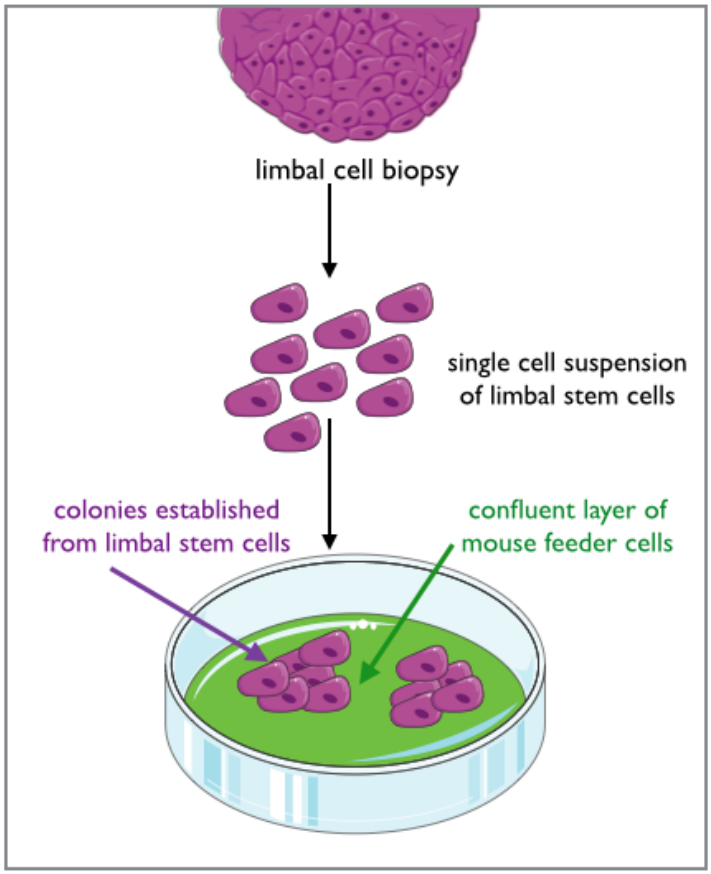


Figure 1. 7.1 Stem cell co-cultivation with mouse feeder layers: the principle

The limbal stem cells, taken from autologous or cadaveric corneal biopsies, are enzymatically cleaved to a single cell suspension and then are plated on top of previously grown and inactivated mouse feeder layers to expand, whilst maintaining their undifferentiated state. The inactivation of feeder cell MEFs is performed either via Mitomycin C treatment or γ-irradiation and has the purpose of arresting the metabolic growth of the mouse cells to prevent their overcrowding in favour of the stem cell culture. (Image copyright author's own).

1.7.2 MLV virus infectivity for human stem cells

The practice of human stem cell cultivation on MEF feeder layers for the purposes of expansion, either prior to clinical transplantation or for biomedical experimentation purposes poses questions in relation to MLV virus transmission. The presence of XpR-1 receptors, necessary for xenotropic MLV infection in human stem cells is long established (Amit et al, 2005 ; Sakuma et al, 2010) and XMRV has been shown to adequately infect human neuronal progenitor stem cells (Ravichandran et al, 2011).

One study has assessed the transmission of MLVs to human embryonic stem cells (Amit et al, 2005). The authors conclude that MEF feeders are not able to pass MLVs onto human stem cells. However, several points have not been addressed in the above study as we will explain below.

Overall, Amit et al, 2005 assessed the transmission of MLVs from MEF cells and from a positive control, an x-MLV-expressing mink cell line, onto human embryonic stem cell lines (hESCs) by co-cultivating these and passaging serially. When the hESCs were cultivated for one passage on the feeder cells, they did not test positive for MLV *pol* RNA after contact with the MEFs but were found positive after cultivation with positive virus-producing control (mink cell line expressing x-MLV). To establish the presence of MLVs in the original MEF cells and the co-cultivated hESCs, the authors measured xenotropic and polytropic MLV *pol* mRNA transcription, *env* gene expression and RT activity. However, the MEF layers were not assessed prior for expression of MLV replication-competent viruses. The RNA MLV (*env* and *pol*) detection through RT synthesis, is not definitive proof of a replicative MLV, as defective or endogenous MLVs may also be present or expressed partially i.e. mutant *gagpol* or *env* without full virion formation may occur.

This obstacle may have been overcome by direct infectivity testing, demonstration of cytopathic effect or syncytia or demonstration of viral particles. The inability to control for replication-competent virus, could be the reason why the study found that hESCs cells were only infected by the cell line that was producing a known x-MLV

virus (positive control) and not from the *pol* or *env* expressing MEF feeder cells. To add, the x-MLV virus from the positive feeder cell line is not described. The authors provide information that reverse transcriptase is produced in the MEF supernatants, but fail to include a positive control from a cell line that produces a replication competent retrovirus, such as their xMLV producing mink cell line for comparison.

Lastly, the authors undertake further infectivity assessment by re-passaging co-cultivated hESCs onto human feeder cells. The mink-cell line infected hESCs are "MLV provirus positive and a robust MLV mRNA expression is observed after up to eight passages on human feeders". The MLV provirus presence is, however, not experimentally proven via integration PCR. In absolute contrast to the article's title, the authors have demonstrated that human stem cells can be infected by co-cultivating mouse layers expressing x-MLV, albeit the mink cell line and not the MEF, and even so at one single passage; a very short time frame compared to the co-cultivation practices of stem cells in clinical practice (which is at least 4 passages or 52 days).

Ecotropic MLV can infect human cells expressing the ecotropic receptor mCAT-1 (Albritton et al., 1989). Recently, Voelkel et al, 2010 showed that human HT1080 cells lacking the mCAT-1 receptor can efficiently endocytose e-MLV irrespective of mCAT-1 presence, as assessed by immunofluorescence of matrix protein MA-GFP and flow cytometry using an anti-matrix protein MA-GFP. The data suggested that a non-specific internalisation pathway exists for e-MLV in human cells that is functioning independently of the cognate receptor. This had been previously suggested by earlier reports over PH-dependent or receptor-mediated viral entry. The demonstration of PH-dependency relies heavily on the inhibition of infectivity via raising the PH of endocytic pathways, that is complete for viruses, such as VZV whereas e-MLVs have been found only partially (80%) replication inhibited (Andersen, 1983, McClure 1990, Nussbaum 1993). Thus, it is implied that a remaining 20% of e-MLV is able to gain entry via independent endocytosis. Such receptor-independent entry is facilitated by the use of cationic liposomes, such as lipofectamine, that can mediate ecotropic MLV infection in the absence of the specific receptor, albeit viral infection is 100-fold times less effective than in the presence of the receptor (Innes et al, 1990).

More importantly it has also been shown (Barnett et al, 2001) that the receptor requirement is abolished in the presence of receptor-compatible soluble envelope RBD complexes or secondary viral infections that facilitate the conformational changes necessary for virus entry (a phenomenon contrary to the retroviral interference principle). This transactivation through a secondary viral release occurs for ecotropic MLV entry through a non-ecotropic receptor; this is likely to occur in the presence of a polytropic MLV virus that is often activated in the presence of e-MLV (Evans et al, 2009). Replicating MCF viruses infect cells lacking the MCF-specific receptor (Syg1) in the presence of trans-activating e-MLV envelope receptor binding complexes or domains (RBD) (Wensel et al, 2003). This has also been shown for Porcine Endogenous Retroviruses (PERV) where transactivation by Gibbon Ape Leukaemia virus or GALV -RBD enabled wild-type or H10A mutant PERVs of all three host-range groups to efficiently infect cells from humans and rodents that lack functional PERV receptors and it substantially enhanced infectivities of wild-type PERVs, even for cells with PERV receptors (Lavilette, 2004).

In Teich et al, 1977 the authors study murine embryonal teratocarcinoma stem cells for their sensitivity to infection by MLVs and find that the murine stem cells are refractory to infection by MLV, unlike non-stem murine cells, but lose this resistance once they differentiate. Proviral MLV DNA integrates but the DNA is transcriptionally silenced. This study is a hint to the notion, that the control of MLV infectivity for stem cells is also a function of the stage of differentiation. This restriction is now known to be due to the transcriptional co-repressor factor, TRIM28 binding on the conserved primer binding site (PBS) in the undifferentiated ES cells. More recently, a stem-cell specific zinc protein ZFP809, has been shown to be responsible for the silencing of MLV expression by recruiting TRIM28 to the PBS on the integrated provirus, when expressed in differentiated cells (Wolf et al, 2015).

Cytosine deaminases, such as APOBEC are important for the epigenetic remodelling in human stem cells (Roa et al, 2010) and as such, they are expected to restrict viral propagation and MLV infection, in concert with toll-like receptors (TLRs), Interferon and NFkB pathways that are stringent stem-cell innate antiviral responses. However,

several human stem cells, such as endothelial progenitor cells, have less APOBEC and are more susceptible to HIV-1 infection (Teofili et al, 2010). Indeed, the antiretroviral action of APOBEC 3G depends on its abundance.

To conclude, although xenotropic MLVs are very well known for their ability to infect human cells and human stem cells, ecotropic and polytropic MLVs such as MCFs are traditionally considered unable to enter human cells lacking the appropriate receptor. A combination of receptor interference however, and the viral infectivity particularities for stem cels as outlined above via either less APOBEC or a differentiation in staging are factors that may influence towards a different intracellular viral fate.

1.7.3 Clinical considerations over mouse and human cell co-cultivation

Cultivation of human keratinocytes and corneal epithelial cells on feeder layers of growth- inhibited mouse fibroblasts in FBS-supplemented medium is the standard culture technique to obtain adequate numbers of keratinocytes for skin grafts and corneal epithelial cells for corneal reconstruction (Pellegrini et al, 1991) . Limbal stem cells that are the focus of this thesis, are located in the niche between the sclera and the cornea at the corneo-scleral junction. Their action is to rejuvenate the surface of the cornea with new cells and also act as barriers to the migration of conjunctival epithelial stem cells to the surface of the cornea. The differentiation cycle and number of stem cells is tightly regulated and is severely decreased in corneal trauma, injury and limbal stem cell deficiencies (Dua et al, 2000).

The expression of restriction factors for viral infection in epidermal or limbal stem cells and their susceptibility to xenotropic gammaretroviruses has not been described. The infection of macaques by XMRV *in vivo*, has demonstrated that a transient low level viraemia may be observed and virus found in all tissues, although an antibody response is not significant (Onlamoon et al, 2011). Although very rare, viral infections of the eye structures are quite able to spread in the bloodstream (Robert et al, 2005) and could constitute a significant transplantation risk. Other human cell prone endogenous retroviruses in the past such as the porcine endogenous retrovirus (PERV), have caused significant concern and attention, hampering transplantation practices

from porcine tissues (Denner et al, 2012). Retroviral integration in human stem cells would pose additional risks, as human endogenous retroviruses HERVH and HERVK seem to play a significant role in the stem cell differentiation process (for a review see Glinksy et al,2015; Wang et al, 2014); and insertional mutagenesis would be likely, mainly due to the *de novo* DNA methylation of infecting proviruses (Cherry et al, 2000).

Johnen et al, 2011 find that human retinal pigment and iris pigment epithelial cells grown on mitotically inactivated MEF feeders are positive for xenogeneic mouse collagen mRNA, derived from the NIH-3T3 feeder layers used for their expansion; a finding that they conclude may limit the use of these cells for transplantation. Indeed, transplantation of mouse foreign material during the practice of limbal stem cell expansion could lead to allograft rejection, immunological reaction and possible viral transmission through the intact MEF fibroblasts as they persist in the stem cell sheet that is transplanted.

To inform the current situation in limbal stem cell transplants, a review (Baylis et al, 2011) identified the outcomes of 28 case reports and case series of a total of 583 patients (597 eyes) receiving cultured limbal epithelial cell therapy that were published over the last 13 years. In this review, cultures were grown using 3T3 cells along with a human amniotic membrane or fibrin substrate during the entire culture process in 16 studies, with the exception of two studies (Schwab et al, 2000) and (Rama et al, 2010), who only used 3T3 cells at the beginning of the culture process and then plated onto a secondary substrate to minimize transplantation risks. Two studies (Pelegrini et al, 1997) and (Daya et al, 2005) had cells grown on 3T3 cells alone. The review identified complications associated with the limbal cell transplant that led to subsequent surgery or a repeat of the limbal graft. In total, 169/597 eyes in 15/28 case studies suffered complications of which 8 were graft rejections (2 of which were complete and 1 was cyclosporin-related). Infectious complications (10) such as infective keratitis (4), infection (2), herpetic keratitis (3), pyogenic granuloma (1), were positively correlated with 3T3 mediated culture expansion (6/16 studies versus 1/9 without 3T3-with FCS and 0/4 animal free cultures).

1.7.4 The CF1 feeder cell-line

CF1-derived mouse embryonic fibroblasts are being supplied by a variety of sources (ATCC, GlobalStem, Applied StemCell, ScienCell, CellSystems, EMD Millipore, Tebu-Bio, Biopioneer, iGentBio, USBio, VHBio, Gentaur -December 2015) as cell lines or primary cells. All CF1 cells that are commercially available are either untreated, Mitomycin C-treated or γ-irradiated. The CF1 cells are premium quality mouse embryonic fibroblasts (MEFs) marketed for optimal sustainment of embryonic stem cell culture in a variety of applications. They are supplied after microbiological testing- free from a variety of mouse pathogens, testing is however not inclusive of investigation for murine leukaemia viruses.

The origin of CF1 feeder cells is quite obscure. CF-1 mice or Carworth Farm-1 are outbred Non Swiss Albino mice whose lineage goes back to wild albino origin, according to several provider declarations. It is believed that the stock is descended from animals bought from a breeder in the 1930s that were originally Non Swiss Albino (NSA). Passed on for distribution to Carworth Farms, Missouri in 1935 and inbred for >20 generations, the stock was outbred from a single pair and released in 1974 by Carworth Farms as CF-1 stock. There are many varieties of CF-1 stock-related mice that are relevant to biomedical research and used invariably in various applications. These are:

CRL:CF1 (Charles River laboratories:CF1). The mice were cesarean derived in 1974 from a “representative cross section of the CF1 stock” by Charles River Laboratories.
Hsd:NSA (CF-1). This is the Non Swiss Albino stock from Harlan Laboratories, that was obtained and developed in turn from an unknown number of CF-1 mice from Charles River laboratories before 1980.
CFLP mice; the stock name of CF-1 was changed to CFLP when Carworth Farms was later renamed to Anglia Laboratory Animals Limited.
CS-1 mice; these are also considered to be a renamed CF-1 stock by the Scientific Products Farm, Manston, Kent, UK.

MF-1 mice; The CS-1 stock is thought to have been crossbred with LACA stock mice (renamed CFW: Swiss Webster mice from Carworth Farms) to produce the MF-1 stock by Oxfordshire Laboratory Animal Colonies, UK (OLAC).
OF-1 mice; Oncine France 1 mice are also considered to be derived from the CF-1 outbred stock and renamed when introduced in France by Charles River in 1966.

In figure 1.7.2, a literature research over the origins and crosses of CF1 cells in global mouse farms and laboratories is attempted. CF1 cells are distinguished from CFW mice (Carworth Farms White) which is a Swiss Webster stock from the Rockefeller Institute, New York inbred for 21 generations by Carnochan and outbred as a closed colony from then on; a stock distributed from Charles River Laboratories since 1974. CFW is identical to the LACA strain (Laboratory Animal Centre Albino) available through Laboratory Animals Centre, Carshalton, Surrey, UK as a renamed CFW since 1970. The LACA stock (renamed CFW) was sent to OLAC (Oxfordshire Laboratory Animal Colonies, UK) before 1970, where it was crossed with CS-1 mice (possibly a renamed CF-1 stock) from Scientific Products Farm, Manston, Kent, UK to create what are now marketed as MF-1 mice and feeder cells.

But are CF1 mice really outbred? Hideo Goto, a Japanese researcher from the National Institute of Agrobiological Sciences, Japan has reportedly inbred the CF1 outbred strain for over 20 generations creating the strain CF1/Sgn (RBRC No 01674) in 1978. The CF1/Sgn strain is an animal model used for cleft palate studies and diabetes induction via alloxan administration. However, this CF1/Sgn strain is also marketed as CF1 and maintained by Jackson laboratories. This adds to significant confusion between the outbred and inbred origin of CF1 mice and related products depending on the originating provider. There is significant reason to indicate that the CF1 mice from which the ATCC CF1 feeder cells are derived, have characteristics suggesting a Japanese *mus mollosinus* crossbreed.

Figure 1.7.4 The origin and current market supply chain of the CF1 stock

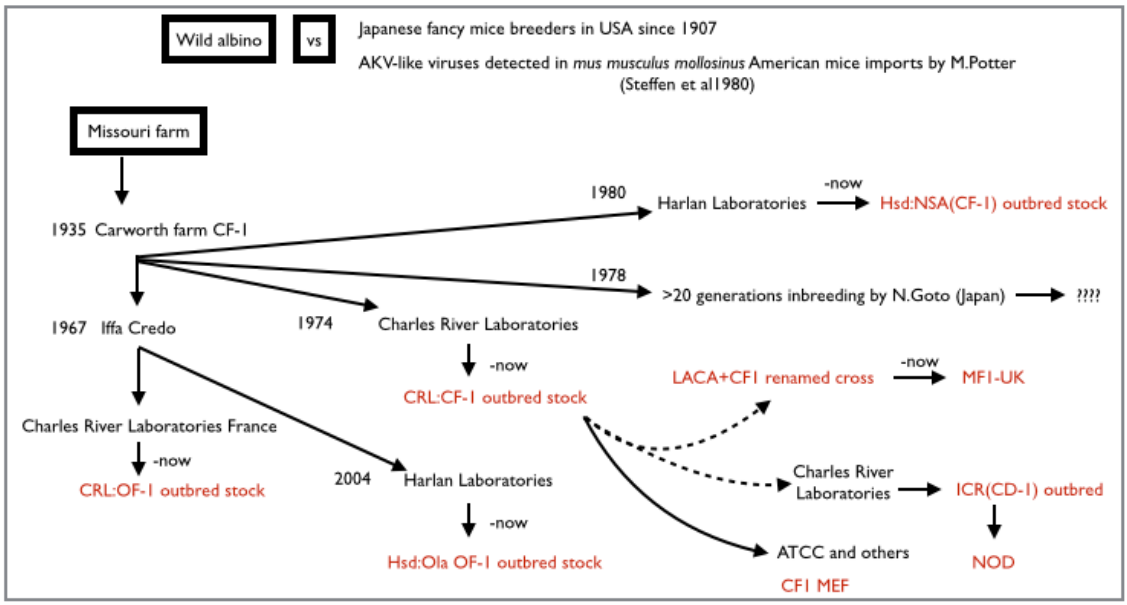


Figure 1.7.4 The origin and current market supply chain of the CF1 stock

The CF1 mouse stock has been traced back to the original reports of its origin in the USA in the 1930s. Reports of several mouse farms obtaining and sourcing the CF1 outbred stock are plotted, while the new names of the stock are shown in red. In black dotted line, two reports of a market cross are shown, where a high suspicion for cross-breed with AKV-bearing mouse strains is sought to exist.

1.7.5 CF1 mice and MLV virus expression

The first documentation of an activated ecotropic MLV virus in CF-1 mice comes from 1975. Reilly and Finkel et al, 1975 first isolated an osteoma-inducing ecotropic MLV virus from benign osteomas of CF1 mice and named it RFB-MLV (after the names of the authors Reilly- Finkel- Biskis). The RFB-MLV ecotropic virus was discovered along with two osteosarcoma inducing viruses, namely FBJ-MSV from a spontaneous osteosarcoma and FBR-MSV from a radiation-induced osteosarcoma; both in CF1 mice. Interestingly, in the radiation-induced osteosarcoma mouse tissue another ecotropic helper virus, FBR-MLV is mentioned by the original authors.

The CF1-related benign RFB-MLV virus was subsequently cloned in 1992 by Petersen (Petersen, 1992). A virus stock of RFB-MLV is made available by ATCC (clone RFB-14, ATCC RFB VR-909). The full genome of the RFB virus however was not characterised until 1996 by Gimbel et al, who undertook the first full molecular and pathogenic characterisation of the isolate from ATCC (Gimbel et al, 1996). It was shown that the RFB virus is present in CF-1 mouse derived cells and is a simple ecotropic retrovirus with no additional oncogene or cellular sequences, that bears significant similarity (>97%) with AKV-MLV λ623, the most pathogenic of the leukaemogenic MLV viruses. The sarcoma viruses, contrary to the ecotropic RFB (that was shown to have the simple retrovirus anatomy ie 5LTR'-*gag -pro- pol- env*- 3LTR') are shown to include cellular oncogene sequences and in particular *v-fos* for FBJ-MSV and a fusion of *gag- fos* gene for FBR-MSV. The other ecotropic virus, FBJ-MLV virus was also described in CF1-related sarcomas, as part of the FBJ-MSV osteosarcoma complex, acting as the ecotropic helper virus for the sarcoma virus stock. The (viral) *v-fos* gene (3.4Kb) is since considered to be this virus' transforming gene for oncogenicity and osteosarcoma formation.

The pathogenicity of RFB-MLV *in vivo* was demonstrated via inoculation on NMRI-mice which lack activated ecotropic viruses (Leib-Mosch et al., 1986). NMRI mice developed lymphomas with an incidence of 91% (mean latency 235 days) and skeletal lesions, with an incidence of 14% versus an incidence of 18% of lymphomas in mock-infected control mice for 408 days' latency. Interestingly, 100% of infected CBA mice

developed osteomas. The authors also described the lack of the generation of MCF-type recombinant viruses in the course of RFB-induced osteoma and lymphoma genesis. However, in osteosarcomas induced in CF1 mice, MCF-type recombinants have been found abundantly integrated in gDNA in clonal fashion (Van der Rauwelaert et al, 1988).

In vitro infectivity in Leib-Mosch et al., 1986 showed that the RFB-MLV virus can infect all types of mouse strains in cell culture, but not the cell line X/Gf; while it distinctively does not produce any cytopathic effect in any mouse cells tested; neither on NIH-3T3 nor Balb/c cells. It was found unable to infect hamster, rabbit, cat, mink, dog, sheep, cow or human origin cells. However, it is able to induce plaques in XC cells, originating from *rattus norvegicus* .

1.8 Thesis aims and outline

The main aim of the present studies is the assessment using modern molecular tools over the low risk of transmission of replication-competent murine retroviruses in human stem cells. To do this we have set out the following experimental objectives:

Aim 1: To develop and optimise the molecular methods required to detect mouse endogenous retroviruses and replication-competent isolates producing reverse transcriptase in cell culture material.

Aim 2: To assess the principal mouse cell lines used in stem cell research as feeder layers, for the presence or activation of murine endogenous retroviruses.

Aim 3: To map the integration of XMRV and MLV viruses in genomic DNA using next generation sequencing.

Aim 4: To co-cultivate mouse feeder and human stem cells by concise replication of human clinical transplantation protocols and assess MLV potential of integration in human stem cells.

Chapter 2
Materials and Methods

2.1 Cell lines and maintenance

22rv1 (ATCC CRL-2505) is a human prostate carcinoma cell line derived from an androgen-dependent CWR22 xenograft, serially propagated in nude mice after castration-induced regression (Paprotka et al, 2010).

LnCap (ATCC CRL-1740) are androgen-sensitive prostatic cancer cells derived from a supraclavicular lymph node metastasis (Horoszewicz et al, 1983).

DERSE (Detectors of Exogenous Retroviral Sequence Elements) cells are LnCap cells, transfected with an iGFP PBABE-Puro construct plasmid. The cells were donated by Dr Lee Kyeongeun, HIV Biology Unit, Centre for Cancer Research National Cancer Institute, Frederick Maryland (NCI, USA). The DERSE cells are able to fluoresce as positive iGFP (induced green fluorescent protein) when infected with a replication-competent retrovirus. Similar luciferase expressing DERSE cells experiments have been published (Aloia et al, 2012).

293T cells (ATCC CRL-3216) are human embryonic kidney cells transformed in 1970 with adenovirus 5 and were donated by Dr O.Erlwein, Imperial College (Pear et al, 1993).

All the above human cell lines were cultured in complete RPMI-1640 medium supplemented with 10% foetal bovine serum (FBS) (Gibco 16000-044) (heat inactivated at 56^{o} for 1 hour) and 1% penicillin G and streptomycin (Sigma, UK).

HCEC or Human Corneal Progenitor cells (Invitrogen, C-018-5C) are human primary cells isolated from a stem cell-rich area of the corneal surface of a normal human eye (limbus). They are cultured in a designated Keratinocyte SFM (Gibco, UK) growth medium in monoculture (Li et al, 2007).

Co-cultivated human and mouse cells were cultured in a supplemented hormonal endothelial medium (SHEM) (Shortt et al, 2005). The reconstituted medium contained Dulbecco /F12 2:1 (Invitrogen cat# 11330-057), FBS 10% (Gibco 16000-044) (added prior to use), Insulin 5 μg/ml (Sigma), Hydrocortisone 0.4 mg/ml (Sigma), Cholera

toxin subunit 0.1nM (Sigma), Penicillin-Streptomycin 50IU/ml and EGF (endothelial growth factor) 10 ng/ml (added at first passage 3 days post-plating) (Sigma).

MEF-J2 (UK Stem cell bank- Altrika Ltd) are murine embryonic fibroblasts (MEF) from Swiss albino mice cells. They were originally developed by Howard Green (Green and Todaro, 1963).

MEF-CF1 (ATCC, SCRC-1040) are murine embryonic fibroblasts from CF1 outbred mice. Cells were obtained from ATCC (SCRC-1040, SCRC-1040.1) Applied Stem Cell (ASF-1225) and Global Stem (GSC-6001).

BALB-K (ATCC CCL-163.3) are murine fibroblast cells derived from Balb-c mice transformed by the Kirsten Murine Sarcoma Virus (Ki-MSV) (Aaronson and Todaro, 1971). The latter is a defective Murine Leukaemia Virus recombinant that only replicates in the presence of a helper ecotropic MLV (Khan et al,2001).

McCoy (ATCC CRL-1696) are murine fibroblast cells, donated by Dr. Steven Kaye, Imperial College that express a high number of murine Intracisternal Type A elements (IAP) (Pomerat et al, 1957).

NIH-3T3 (ECAAC-93061524) are murine embryonic fibroblasts established from a NIH Swiss mouse embryo as a continuous cell line and used in the assessment of syncytia formation by murine leukaemia viruses (Jainchill et al, 1969).

XC cells (ATCC CCL-165) are derived from a tumor of *ratus norvegicus* induced with Rous Sarcoma Virus (RSV) (Ogura et al, 1976).

Mus dunni (PHE- 94101211) –kindly donated by Prof. Jonathan Stoye, NIMR, London- are mouse tail fibroblasts, that are easily infectable by 4 classes of MLV viruses (ecotropic, amphotropic, xenotropic and mink cell focus-forming viruses) and are used as an amplification stage for MLV viruses prior to infectivity assay or PERT assay determination (Lander et al, 1984).

All murine cell lines were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS and 1% Pen/Strep. Cells were passaged when 80-90% confluent. Adherent cell monolayers were first washed with phosphate buffered saline (PBS), then incubated with Trypsin- Ethylen-diamine-tetraacetic acid (EDTA) at 37^{o} for 5 min. Cells were then resuspended in growth medium and centrifuged at 400 g for

5 min, then resuspended in complete growth medium. All culture incubation was performed in humidified incubators in 5% CO_2 and 37^oC.

All cells were cryopreserved in liquid nitrogen (LN) in vials containing the appropriate growth medium and 20% FBS, 1% DMSO until needed; they were defrosted at 37^o and centrifuged at 400g for 5 min. The supernatant containing DMSO was decanted and cells were resuspended in growth medium at 37^o in a humidified CO_2 incubator. Cell lines are summarised in Table 2.1. Transfection of *mus dunni* cells with viral inoculates was preceded by the use of Polybrene (TR1003G- Millipore) to improve transfection efficiency at least 10-fold (Kawai et al, 1984).

Table 2.1 Cell lines used in the thesis

Cell culture				
species	type	name	description	reference
human	Cell lines	22rv1	prostate cancer cells	Paprotka et al,2010
		LnCap	prostate cancer cells	Horoszewicz et al, 1983
		LnCap-DERSE	prostate cancer cells with reporter plasmid	Aloia et al, 2012
		293T	human embryonic kidney cells	Pear et al, 1993
	Primary cells	HCEC	human corneal progenitor cells	Li et al, 2007
mouse	Cell lines	MEF-J2	MEFs used in clinical limbal transplantation	Green and Todaro,1981
		MEF-CF1	MEFs used in research stem cell applications	Petersen et al, 1993
		Balb-K	Ki-MSV transformed balb-c cells	Aaronson and Todaro, 1971
		XC	ratus norvegicus cells	Ogura et al,1976
		McCoy	MEFs expressing IAP particles	Pomerat et al, 1957
		NIH-3T3	MEFs from NIH Swiss mice	Jainchill et al, 1969
		Mus dunni	Mouse tail fibroblasts	Lander et al, 1984

2.2 Human stem cell and mouse feeder cell co-cultivation

The protocol used in human and mouse co-cultivations was adapted from Shortt et al, 2005. In detail the process includes:

I. Establishment of CF1 MEF monolayers

One million cells from frozen are plated per 10 cm plate (~170,000 per well of a 6-well plate). MEFs are seeded and incubated for a minimum of 8h before used.

II. Growth arrest of MEFs with Mitomycin C

MEF cells are growth arrested after 48h by incubating them in medium containing Mitomycin C (Sigma) at a concentration of 4μg /ml for 2h.

III. Co-cultivation

HCEP cells are maintained in Keratinocyte-SFM (1X) medium. After detachment is performed using Accutase (Invitrogen,UK), stem cells are transferred on plates previously plated with MEFs. Briefly, $6x10^3$ stem cells/cm^2 are added to MEF growing cells, seeded at $2.2x10^4$ cells per 6-well plate and allowed to settle within 24 hours. The medium is changed every 2 days.

IV. Passage of stem cells grown on mouse feeder layers

Visibly differentiated colonies (aberrant keratinocyte-like growth, irregular shape) are removed manually under a microscope in sterile conditions via slow-vacuum aspiration or pipette tip scraping. Cells are then washed with either warm supplemented hormonal (SHEM) medium or PBS. The incubation at 37ºC for 5-10 minutes with 1ml of Collagenase IV (Invitrogen, UK) per well results in a visible curling of the stem cell colonies around the edges. Using a cell lifter (i.e. Corning # 3008) the entire well is scraped to remove the colonies.

The supernatant containing stem and mouse cells is transferred into a conical tube. By mild centrifugation (2,000g per 2 min) and trituration medium-small fragments of the original stem cell colonies (~50-200 cells per fragment) are obtained. 1 ml of the pellet

is then replated onto a well of a 6-well plate of prepared MEFs pre-washed with PBS and containing 1 ml of SHEM medium. Cells are then incubated and monitored daily for growth. Cells are passaged after 14 days of culture and re-seeded at a density of 1 × 10^3 cells/cm^2 on a MEF feeder layer. They are then cultured for another 14 days (passage 2-3 and 4).

V. Adherence plating

Using sequential adherence plating it becomes possible to remove contaminating MEFs from the desired human stem cell output, as the MEFs easily and rapidly adhere onto plastic surfaces compared to human stem cells. Hence, harvested mixed cells (HCEP/MEF) are left to adhere on 6-well Nunc plastic plates for 30min in 1ml of SHEM medium. The mouse cells in the mixture are the first to adhere to the bottom of the well, and this provides an opportunity to isolate the human stem cells floating in the supernatant. This procedure is repeated twice for optimal isolation.

2.3 Chemical stimulation of endogenous retroviruses in mouse cells

Induction treatments of mouse cell lines were performed after growing an initial 5x10^4 seeded MEF cells in 6-well plastic plates (Nunc). Chemical stimulants 5-azacytidine-C (5-Aza-C) (Sigma, A2385) and 5-iododeoxyuridine (5-IDU) (Sigma, I7125) were reconstituted according to manufacturer's instructions and stored in 4^o and -20^o respectively. Treatment was carried out in three concentrations for each stimulant, except for Balb-K cells where optimised quantities for its use as positive control have already been published (Khan et al, 2001). Confluent cells were washed and new medium was replenished before an appropriate quantity of the chemical stimulant was added drop wise in the medium and cells were incubated and assayed at regular times. All experiments were assayed in triplicate. Non-treated cells were used as negative controls.

2.4 Infectivity assays

2.4.1 Syncytium assay on NIH-3T3 cells

Confluent NIH-3T3 cells, grown in 6-well plates for 48 hours were inoculated with virus suspensions (300-900µl) arising from centrifuged supernatant from growing MEF mouse cells. In detail, putative virus containing supernatant was harvested from $1x10^6$ cells (CF1, J2 and Balb-K) grown in 6-well plates in 4 ml of total growth medium. The presence of syncytia was assessed daily. Two wells with growing uninfected NIH-3T3 or XC cells served as negative controls. Ten-fold serial dilutions of the inoculant supernatant were also performed. In detail, $1:1\text{-}10^{-4}$ of the original inoculate volume (300-900 µl) originating from 4ml of supernatant of growing confluent cells, were dropped into 2ml of growth medium over NIH-3T3 or XC cells. Supernatant from infected cells was passed serially post 7 days onto uninfected 3T3 cells to determine viral infectivity.

NIH-3T3 cells were also co-cultivated with putatively virus-expressing MEF cells. In the syncytia assay via co-cultivation, NIH-3T3 cells ($3x10^4$ cells/well) were seeded in the same well with untreated Balb-K cells (10^5 cells/well) and with untreated CF1 (10^5 cells/well) in duplicate in 6-well plates.

The presence of multinucleate syncytia due to the fusion of virally infected cells was monitored over a period of 7 days by phase contrast microscopy. A scale for relative syncytia formation was based on McClure et al, 1990: (+++) >25% of nuclei in syncytia, (++) 5-25% nuclei in syncytia, (+) 1-5% nuclei in syncytia, (-) no discernible syncytia formation.

2.4.2 XC plaque forming assay

XC cells were used in XC plaque formation assays both via supernatant inoculant (modified XC assay) and via co-cultivation (traditional XC assay) (Klement et al, 1969). The cells were co-cultivated on confluent XC rat cells and plaques formed

were assessed after additionally fixing the co-cultivation at -20^{o}C with methanol and staining with 0.05% Crystal Violet for 1h.

2.5 Molecular Methods

2.5.1 Nucleic acid extraction from cell culture supernatant

Viral nucleic acids were extracted from cryopreserved or freshly harvested supernatant fluid. Where stated, virus was concentrated from 4-35 ml of supernatant mixed with varying quantities of phosphate buffer saline (PBS) up to 38.5 ml in thin tubes (Thinwall, Ultra-Clear™, Beckman Coulter No 344058) by ultracentrifugation at 75,000 g for 1h at 4^{o}C. Supernatants were decanted and invisible viral pellets were resuspended in 90µl of PBS.

Concentrated virus and non-concentrated supernatant were processed with the QIAamp viral RNA extraction kit (Qiagen, No52906) or the VirusSpin Min Elute Kit (Qiagen, No 57704). Samples were lysed with 540µl of buffer AVL (supplied) containing carrier RNA at 10 ng/µl. The method uses a silica-based column that is centrifuged after sequential ethanol precipitations at 16,000 g for 2 min. Absorbed nucleic acids are washed and the column is then dried by further centrifugation at 16,000g for 2 min after which, the elution of nucleic acids is performed, using 30-100µl of AVE buffer (proprietary kit formula).

RNA specific extraction (total RNA) was also carried out with the RNAeasy Mini Kit (Qiagen No. 74104) that elutes 50µl of total RNA from an original sample volume of 700µl. To purify this RNA, DNAse on-column digestion treatment (Qiagen, No79254) was carried out. To homogenise cells derived from cell culture, Qiashredder columns (Qiagen, No79654) were used prior to extraction, using centrifugation as per manufacturer instructions.

Large number of samples from the chemical induction of murine cell lines were subjected to automated nucleic acid extraction in an automated platform (Magna Pure

MP-96) (Roche, MN) using the MagNA Pure 96 DNA and Viral NA Small Volume Kit (Roche, No 06543588001). The automatic extraction method extracts 50µl of viral nucleic acids from an initial sample volume of 200µl. For genomic DNA-specific extraction, the DNAeasy Blood and Tissue Kit (Qiagen, No69504) was used.

2.5.2 Reverse transcription

RT synthesis mix consisted of 2µl molecular grade H_20, 4µl dNTPs 10mM, 4µl $MgCl_2$ at 50mM, 4µl 10xPCR buffer (H_2O 90%, 0.1M TrisHCl, pH 8.8) to a total of 15µl. To this is added 10µl of the viral lysate or viral NA. cDNA synthesis takes place at 37^o for 90 min followed by heat inactivation at 95^o for 10 min. cDNA synthesis cycling conditions were optimised of sensitivity; an example being the duration of the RT heat inactivation step, that was optimised from 5 to 10 min for better cDNA yield.

2.5.3 PCR amplification for MLV gene targets

The Qiagen QuantiTect Probe PCR kit (Qiagen, No204343) was used. The reaction comprised of 12.5µl Master Mix (supplied), 1.95µl molecular grade H_20, 22pM forward primer, 22pM reverse primer and 11pM probe. Two-step nested-PCR was employed for MLV *gag* gene targets. The first round reaction (25µl) consisted of 0.5U of AmpliTaqGold (AB No N8080240, Warrington, UK), 1x Taq Gold reaction buffer, 1.5mM $MgCl_2$, 200mM each dNTP, 2.5 pM MLV_*gag* reverse outer primer, 2.5pM MLV_*gag* forward outer primer and to this, 5µl of test sample DNA were added. Cycling conditions were 1 cycle of 94^o C for 8 minutes, 35 cycles of 94^o for 30 seconds, 55^o C for 30 seconds, 72^oC for 30 seconds and one cycle of 72^oC for 7 minutes. Second round PCR was set up with identical mix and the inner primers MLV_*gag* inner forward and MLV_*gag* inner reverse. One µl of first round PCR product was transferred to the second round mix and PCR performed in identical conditions. The products were visualised on a 1% agarose electrophoresis gel.

A set of primers and dually labeled probe for Real Time PCR (Dr. P. Tuke, NHSBT/ PHE) was directed against XMRV *gag* and several polytropic MLV *gag* genes. The reaction is set up with 12.5µl QuantiTect Probe PCR mix (Qiagen, UK), 1.95 µl H_20 (molecular grade) and 0.55 primer and probe mix. The cycling conditions are 50^o for 30min, 95^o for 15min and 45 cycles of 95^o for 15 seconds and 60^o for 45 seconds, in the Bio-Rad CXR RT-PCR thermocycler. A one step RT-PCR reaction with the inclusion of RT enzyme (1 µl) (QuantiFast RT enzyme, Qiagen) and the same cycling conditions, was evaluated for RNA targets.

This assay was also modified to include a reverse transcription step, using the standard cDNA synthesis protocol: 2 µl DEPC H_20, 4 µl dNTPs 10mM, 4µl $MgCl_2$ 50mM, 4µl 10xPCR buffer (DEPC H_2O 90%, 0.1M TrisHCl, pH 8.8), 30pM XMRV reverse primer and 1µl of QuantiFast RT (Qiagen,UK), an heterodimeric recombinant enzyme expressed in *E.coli*. Ten µl of cDNA is then transferred to 15µl of PCR reaction mix and the result of the amplification is monitored using the thermocycler CXR (Biorad, CT) software.

A positive control for PCR assays was the VP62_ XMRV DNA, a cloned plasmid donated to Imperial College by Dr. R. Silverman (full XMRV cloned virus) or XMRV derived from growing 22rv1 cell supernatant. For human derived samples, Human DNA (Applied Biosystems, No1202082) was used as a reference control in serial dilutions and for mouse, DNA was manually extracted from maintained MEF cell lines. A quantification curve for real time PCR for MLV *gag* can be seen in Figure 2.5.3.

Figure 2.5.3 A quantification curve in Real Time RT-PCR for MLV *gag*

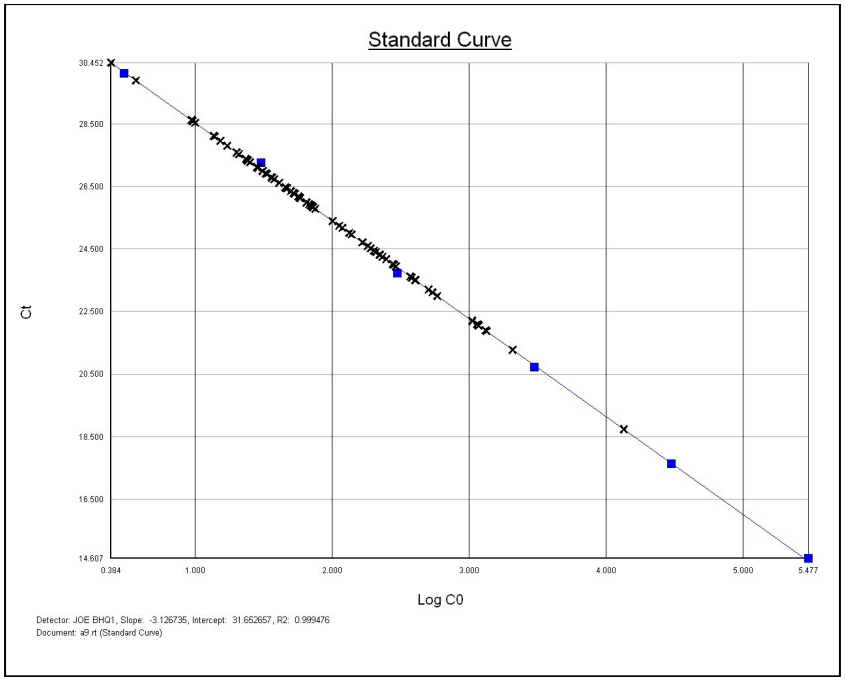


Figure 2.5.3 A quantification curve in Real Time RT-PCR for MLV *gag*

Quantification curve for the Reverse transcriptase one-step Taq-PCR for MLV *gag* was performed with a standard cDNA derived from 22rv1 cells producing XMRV virus in 10- fold serial dilutions (square blue dots). Tested samples' CT amplification (black x dots) is seen on the standard curve (square blue dots) and this allows for their quantification on the y axis. The curve shows the dilutions from 1 million to 10 copies (LogC0) of XMRV containing 22rv1 cells per PCR reaction. Slope: 3.22 and R2: 0.99. The output is from the AB7500 Thermocycler (Applied Biosystems, UK).

2.5.4. Oligonucleotides designed for PCR amplification

MLV *gag* (conserved) primers were directed against a conserved region of MLV *gag*, able to amplify MLVs from an array of different mouse strains (MLV-*gag* conserved) (Dr S.Kaye- Imperial College). The primers were optimised in a two-step nested-PCR format. Parallel amplification on mouse DNA extracted from MEF and McCoy mouse cells served as positive controls. A dually labelled fluorescent probe was also designed to optimise the assay for Real Time PCR. A set of primers and probe for real time PCR (XMRV/pMLV *gag* MLV-Tuke) was directed specifically against XMRV *gag* and spanning for several Polytropic MLV *gag* (Tuke et al, 2011).

A set of primers on relatively conserved regions of the MLV envelope (MLV *env*) was developed using Primer 3, producing a product of variable length size according to MLV isolate (range 600-750bp) due to envelope gene variations. MLV *env* primers were optimised for conventional PCR and downstream direct sequencing. A primer set (MLV mRNA *env*) that spans the exon of the spliced *envelope* mRNA, i.e. lies on both sides of the spliced coding sequence exon for the messenger RNA transcription of MLV *envelope*, was also designed to allow for expression detection of spliced *envelope* mRNA signals.

To differentiate within ecotropic MLV of the AKV family we designed a primer set (MLV tandem) that spans a particular sequence of multiple tandem- repeat containing LTRs. These LTR-specific primers are directed against the LTR of AKV-MLV and in particular the tandem repeats, the junctions of which are found unique to AKV-MLV.

For the splinkerette assay, we optimised the primers found in Adams et al, 2009 that were directed against Mo-MLV. For CF1 directed targets we developed the CF1 splinkerette primers (inner and outer) that are found on the RFB-MLV genome in the U3 (As in Adams, 2009 we also aimed at the 5'-LTR and as such the splinkerette primers are reverse 5'-3') and a second set was also specifically developed for the XMRV positive control (ie XMRV splinkerette inner and outer). Adaptors and primers for the adaptors (short and long) were used same as described in Adams et al,2009 .

All the primers used in the thesis are presented in the Appendix 1. Figure 2.1 shows the position of several of the primers directed against MLVs in an alignment of different MLV isolates.

Figure 2.5.5 Position of MLV gene-directed primers on alignment of MLV isolates

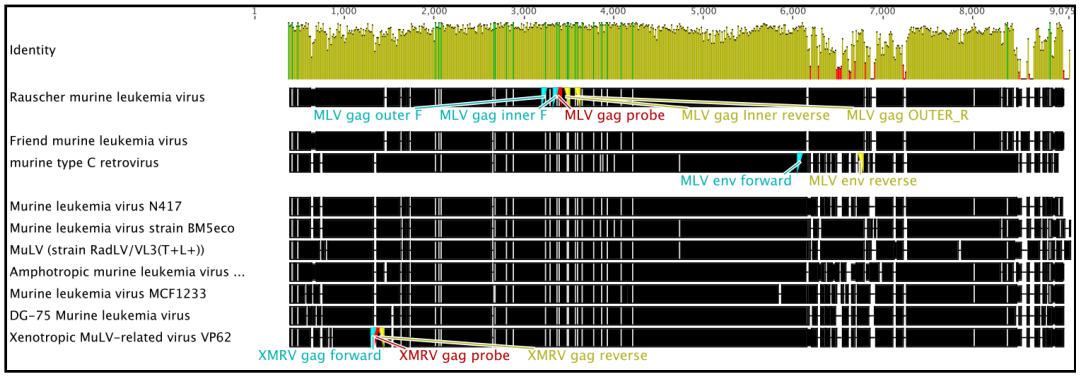


Figure 2.5.5 Position of MLV gene-directed primers on alignment of MLV isolates

Primers were mapped against an alignment of 10 MLV isolates including ecotropic, polytropic and xenotropic isolates (in order of appearance). Forward primers are depicted in blue, reverse primers are depicted in green and the position of probes is depicted in red. Within base pair alignments, black represents regions of high similarity and gaps represent significant differences. The Identity graph at the top of the figure (green wave) demonstrates the conservation of base pairs across the alignment (measured from 1-100). Areas with large troughs (red) are areas of low similarity and higher genetic variability. The graph and alignment was created using Geneious, version 7.0 (Biomatters, MA).

2.5.5 Gel electrophoresis, gel extraction and purification

PCR products were visualised on a 1% agarose gel, using standard electrophoresis under 120V for 1hour. The products were mixed with 1X solution of gel loading dye (GelPilot, Invitrogen). Molecular ladders, Hyperladder I or Hyperladder II (Bioline BIO-33054), were used variably. Denser agarose gels (4% agarose) running for 120V for 2 hours were used for the analysis of multiple PCR products arising from the integration splinkerette assay (see section 2.7 for a detailed protocol).

Purification of PCR products was performed with the Qiaquick PCR purification kit (Qiagen 28104) or Genejet spin columns (Fermentas No K0701 UK) using the buffers supplied by the manufacturer. Each product was mixed with 2 volumes of binding buffer, assessed for PH via a colorimetric indicator. The buffer contains guanidine thiocyanate. Post centrifugation, ethanol washes and drying, the purified DNA is precipitated under low-salt conditions with Tris-EDTA and assessed by nanospectrometry for quality and quantity relevant to downstream applications.

Gel extraction was also performed, by directly cutting bands from electrophoresis gels under UV lighting. The slices were weighed and then mixed with equal weights of binding buffer. Incubation at 56^{o} for 1h allowed the gel to be completely dissolved. Purification of DNA followed the identical process described above for PCR products.

2.5.6 Direct (Sanger) sequencing

Short PCR products (<800bp) were analysed by the direct sequencing service in PHE, Colindale. DNA sequencing for short fragments is performed on Applied Biosystems' capillary sequencers (ABI Genetic Analyser Platforms 3730 and 3130XL), via single read dideoxy Sanger method. Histograms arising from the sequencing result are trimmed at the ends to match primer binding sites, assessed for quality (above 50-70%) and screened for conflicting ambiguities, where an additional second opinion is sought. The analysis of the resulting sequences is performed using the BioEdit

software (BioEdit,MA) or Geneious (Biomatters, MA), and sequences were then submitted to NIMR Gene BLAST databases for further identification.

2.5.7 Product-enhanced reverse transcriptase activity assay

The Product-enhanced reverse transcriptase assay (PERT) is a two-step PCR reaction that detects and quantifies the amount of reverse transcriptase enzyme present in cell culture derived samples. The principle behind PERT is based on a protocol described in Pizzato et al, 2009. It is based upon the action of a putative reverse transcriptase upon an RNA template (MS2 bacteriophage RNA, Roche Diagnostics, MA). The formed cDNA is generated due to the sample's inherent reverse transcriptase activity. The cDNA is then detected via specific template primers and a dually labelled fluorescent probe in a second step, in a real time Taq-PCR reaction (Figure 2.5.7A).

The virus lysis buffer (NZ buffer) consists of 20 mM TrisHCl, pH 7.5, 50 mM KCl, 0.25 mM EDTA, pH 8.0, 0.025% Triton X-100, 50% glycerol and 0.2 mM dithiothreitol (DTT), with the DTT added freshly before use. The method is identical to the one used in Ma et al, 2001. In brief, 10µl of a briefly centrifuged supernatant fluid (harvested freshly or stored at -80^{o}) is lysed with 90µl of in house lysis buffer. The lysate is dissolved on ice for one hour and 10µl are then transferred to a prepared cDNA synthesis reagent mix.

The MS2 template is a bacteriophage RNA virus that replicates without an intermediate DNA replication step. The primers and probes directed at the 2,349bp MS2 genome were chosen from a protocol for the detection of Hepatitis E (PHE, Colindale) where MS2 RNA serves as an internal amplification control. They can be found in Appendix 1.

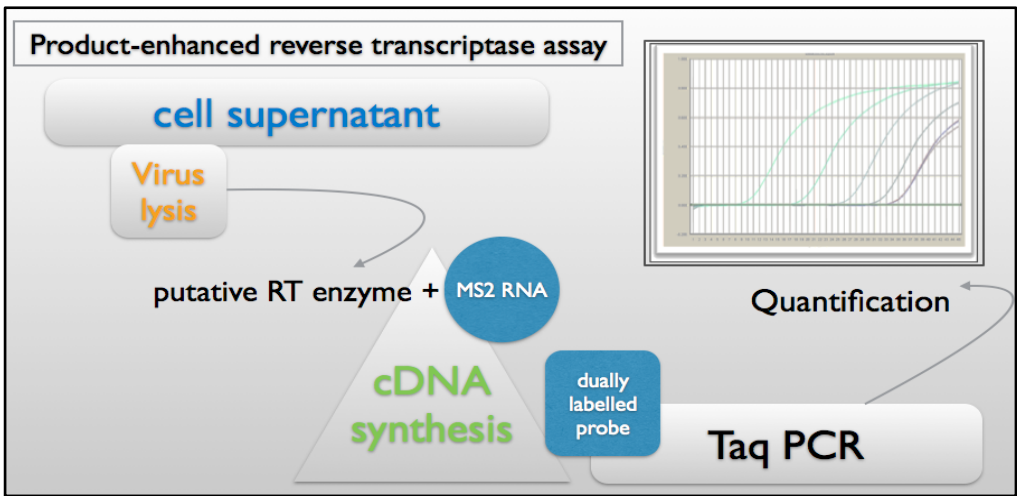
The RT synthesis and PCR amplification were performed as described in Section 2.5b and 2.5c. A standard curve was constructed using the commercially available MLV Reverse transcriptase (Mo-MLV) (Invitrogen, No 28025-013) at 200 iu/ml stock

concentration in ten and/or hundred-fold dilutions after dilution in the lysis buffer (Figure 2.8A). The addition of RNAse A (from bovine pancreas; 10 A260 units [500 ug/ml], Roche catalogue No 11119915001) was also evaluated in various concentrations (see Appendix 2). The lower detection limit of the quantitative real-time PERT was approximately 10^{-10} units of MLV–RT per reaction. The standard curve of MLV-RT covering the range of 10^{-2} to 10^{-8} units had a regression co-efficient R^2 of 0.99, a slope of -3.2 and an amplification efficiency of 99.8%.

A series of optimisation assays were performed to control for low grade background amplification noise in the PERT assay and these are described in detail in Appendix 2.

Figure 2.5.7: Product-Enhanced Reverse Transcriptase (PERT) Assay

A.



B.

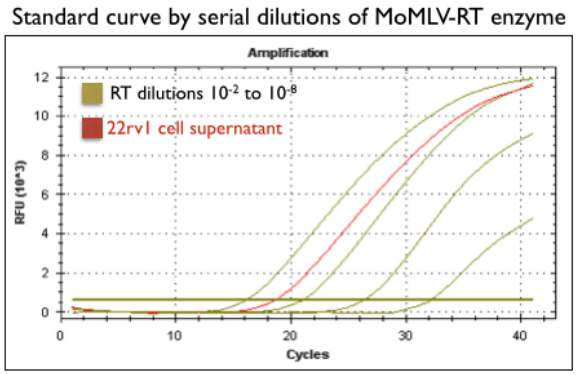


Figure 2.5.7 Product-Enhanced Reverse Transcriptase (PERT) Assay

A. **Description of the principle of the PERT assay:** Cell supernatant is lysed to release viral enzymes, the lysate then acts upon a target RNA template (MS2) in a reverse transcriptase synthesis reaction and the putative cDNA produced is detected by a dual-fluorescent tagged probe in Taq-PCR. A standard enzyme in serial dilutions is necessary for standardisation and quantification of the reverse transcriptase titres (shown as cycle threshold amplification plots in B and C).

B. **Standard curve for the quantification of RT activity**: Four 100-fold dilutions of Moloney-MLV RT enzyme (range 10^{-2}-10^{-8} iu/ml) are used as RT standards for the construction of a quantification curve (Green amplification curves). A cell culture derived amplification curve, also serving as a quantification curve (not shown) is uniquely shown for 22rv1 supernatant (red amplification curve).

2.5.8 Transcription PCR

To measure transcription of target MLVs via spliced mRNA detection, the use of a commercially available mouse HPRT (Hypoxanthine-guanine phospho-ribosyl-transferase) mRNA endogenous control was evaluated (Qiagen, QT00166768). HPRT was chosen as this gene is expressed in very low levels i.e. 1-10 mRNA molecules per cell and thus it is suitable for gene expression evaluation of low copy targets or rare mRNAs. Also, this low copy expression is useful in quantitation of the mRNA signal per cell as the HPRT expression can be considered =1 copy/cell .

The primer set for the envelope transcription PCR (Appendix 1 : mRNA *env* primers and probe) was validated on a Real Time PCR platform (AB 7500, Applied Biosystems, UK) using a prior RT synthesis at 42^{o} for 50 min and inactivation at 70^{o} for 15min with gene specific priming. The PCR mix consisted of 12.5µl QuantiProbe mix (Qiagen), 900nM of both primers and 250nM of the dually labelled fluorescent probe. Serial dilutions of total cell RNA from CF1 cells growing in culture (day 2 post passage), CF1 cells that were stationary (day 5 post passage) and Mitomycin C treated CF1 cells were evaluated.

2.5.9 Integration site mapping via splinkerette PCR

The protocol used for the identification of retroviral insertion sites is derived from Adams, et al 2009 with primer-specific modifications to the target MLV isolates. Splinkerette PCR allows for the amplification of sequences that are found between a known DNA sequence, such as the Mo-MLV viral LTR and a nearby restriction site. Oligonucleotide sequences (adaptors) are designed with overhangs at the 5'-end that are compatible with the overhangs generated by the restriction enzyme Sau3AI on the genomic DNA (Figure 2.4A). The adaptors are ligated at the overhang site and the products are digested with a second restriction enzyme that allows for the removal of unwanted viral LTR-adaptor bi-products. The (viral LTR-genomic insertion site-adaptor) product is then amplified with nested PCR (Figure 2.10).

The steps in the splinkerette reaction are in detail (also shown in Figure 2.10):

Extraction of genomic DNA: gDNA used in the splinkerette reaction is assessed free from salt and organic contaminants in photo spectrometric evaluation for purity. For optimal gDNA eluate, mouse-DNA free columns were used i.e. not likely to contain mouse sequences; such is the Qiamp DNA extraction kit (Qiagen). Freshly extracted and quantified DNA is stored at -20^{o}C.

Digestion of genomic DNA with the first restriction enzyme (Sau3AI): The choice of restriction enzyme used is determined by the respective cutting site on the viral LTR. A safe distance downstream of the viral LTR is needed to allow for amplification of a reasonable length amplification product (minimum 100bp). The restriction enzyme chosen cuts in a relative average frequency human and mouse genomic DNA to ensure amplification efficiency.

Restriction enzymes that leave 4bp overhangs are ideal for maximum adaptor ligation (Adams et al, 2009). From this protocol, I evaluated SAu3AI (NEB, R0169) for its ability to digest MLV viral LTR sequences in appropriate sites on the 3'-end viral LTR (Figure 2.5.9). Enzyme is added at 20IU/reaction, with NEB Buffer 1 (CutSmart-2012) was used at 1x, and BSA at 0.3 µg/µl (NEB, R0169) for use against 2µg of genomic DNA (1µl) in a 30µl reaction and the mix is incubated at 37^{o} for 12-16 hours or overnight. Heat inactivation takes place at 80^{o} for 20 min freshly before the next step, ie the ligation.

Design of the adaptor sequences: The adaptor sequences used were first reported by Mikkers et al, 2002 and are shown in Figure 2.5.9 (and Appendix 1).

Splinkerette adaptor ligation: The two splinkerette adaptors (short and long) are mixed in NEB Buffer 2 (250mM NaCl, 50mM Tris-HCl, 50mM $MgCl_2$ and 1mM dithiothreitol DTT, pH:7.9) then denatured and annealed by heating in 95^{o}C for 5 min and then brought to room temperature at a rate of -1^{o}C/min and kept on ice. The adaptor mix (1µl per reaction) is T4-ligated with the Sau3AI digested gDNA (4.5µl

from the digestion reaction). The T4 ligation reaction uses T4 ligase (NEB, M0202S) (20IU), T4 ligase buffer (1x) and molecular grade H_2O up to a volume of 100 µl/ reaction. Incubation is at 4^o overnight, heat inactivation takes place the following day at 60^o for 20 min and the ligation reaction product can be stored at -20^o indefinitely.

Second digestion: As the viral LTR sequence for all target MLVs is repeated internally, on a second step, the splinkerette products from the ligation reaction are re-digested to prevent the amplification of internal fragments. In the case of primary SAu3A1 restriction, EcoRV is recommended (Adams et al, 2009).

Clean up: The splinkerette adaptor-ligated and EcoRV-restricted DNA is then purified with an appropriate kit (Qiamp PCR purification kit) prior amplification.

Two-step nested PCR amplification: A proofreading (pfu) polymerase is used in the first step while the second step is used to increase the specificity of isolating virus/ genome junctions. The products are then visualised on a 4% agarose gel and purified for sequencing.

In brief, 300ng of digested, ligated and purified gDNA is used in a first PCR reaction with pfu polymerase as follows: Pfu polymerase (Stratagene, 600320) (1.25U), Pfu 10x Buffer (1x), dNTPs (Sigma, D7295) (0.25mM), LTR primer (10µM), splink1 primer (10µM) and molecular grade water up to 50µl total reaction volume. The conditions are 94^o C, 3 min, 94 ^{o}C, 15 seconds, 58 ^{o}C, 30 seconds, 72 ^{o}C 3 minutes – 29 cycles 72 ^{o}C, 5 minutes – 1cycle 40^oC – hold.

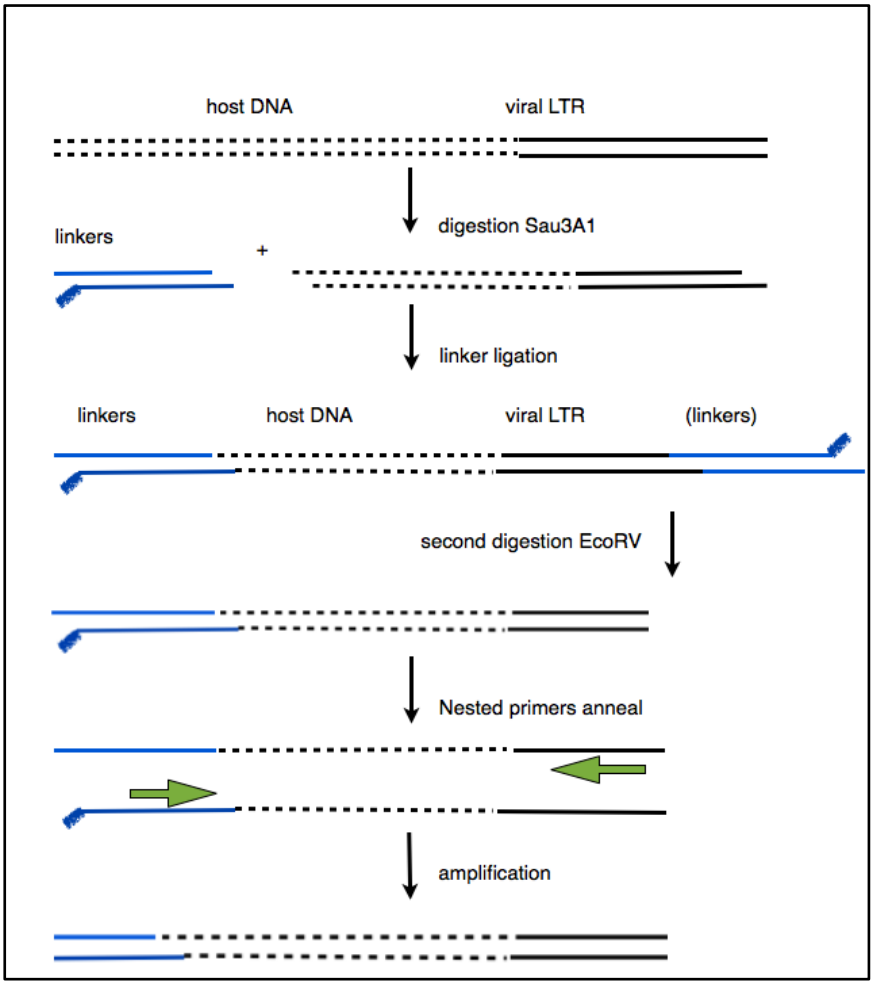
The second reaction is set up with Qiagen Multiplex PCR (Qiagen, No206143) mastermix (25µl), second step (inner) primers at 10µM and the primary PCR product (1µl from first step). The reaction is brought to 50µl with 22µl molecular grade water and is set to cycle in the conditions: 94^o C 15min (prolonged as the Qiagen Multiplex Master Mix has a longer activation time) 94^o C, 15 seconds, 58^o C, 30 seconds, 72^o C 5 minutes – 25cycles, 72^o C, 5minutes – 1cycle, 40 ^{o}C – hold.

Gel electrophoresis: The products are visualised on a 4% agarose gel, and will discriminate fainter bands likely to occur in multiple integrations. The gel runs at 120V for 1h 30min.

Sequencing: After purification, direct Sanger sequencing is carried out at a dedicated facility (PHE genomic services, Colindale) and sequences are analysed with Geneious, version 7 (Biomatters, MA) and identified using GenBank and BLASTn databases.

Figure 2.5.9: The splinkerette reaction principle

A.



B.

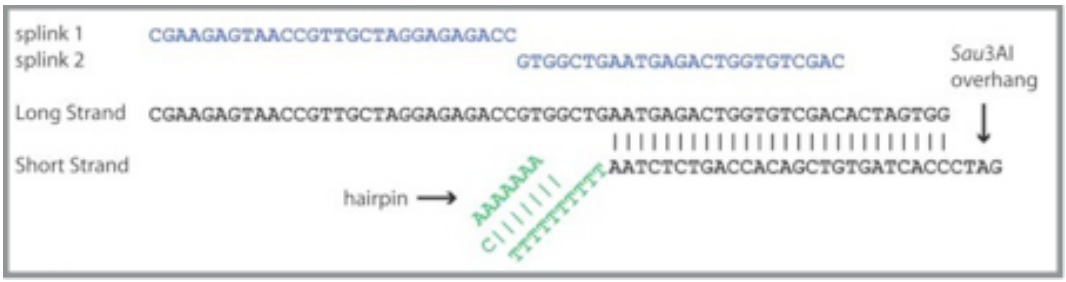


Figure 2.5.9 The splinkerette reaction principle

A. Splinkerette PCR assay principle: Genomic DNA is digested with Sau3A1, a 6-base pair cutter enzyme that leaves an overhang; a predesigned oligonucleotide adaptor ligates on this overhang site while a second restriction enzyme cleaves ligated adaptors on pre-determined viral LTR positions. Nested PCR amplification of the ligated products produces a product that contains at least 50bp of genomic host DNA sequence. Notably the primers and restriction enzyme positions are designed to target 5'-LTR genomic junctions.

B. The splinkerette adaptor sequences in 2-D confrontation: the two adaptors are one long strand and one short strand as shown in the figure. In green is the base pair sequence arranged in a secondary hairpin structure; this is introduced to prevent the formation of self dimerisation products and splinkerette owes its name to this particular design of the short strand adaptor. In blue are the primers targeting the splinkerette adaptors in the final nested PCR step (splink 1 and splink 2-found in Appendix 1).

2.5.10 Next generation sequencing

The next generation library construction for viral nucleic acids in tested cell culture eluate and for PCR products post amplification was performed using two manufacturer supplied library preparation kits. The TrueSeq Stranded total RNA with Ribo-Zero Gold (Cat RS-122-2201) and the Nextera DNA library preparation kit (Cat FC-121-1031) (both Illumina, Cambridge) were evaluated and used. Samples were prepared according to manufacturer's instructions, library pools were assessed for quality and quantity using a 2100 expert High Sensitivity DNA and RNA MicroChip assay (Agilent, UK) and were submitted to next generation sequencing on a MiSeq Illlumina platform (Hammersmith Campus, Imperial College).

Bioinformatic analysis of obtained sequences was analysed with Geneious version R7 and R8 (Biomatters, US) and analysis of protein function effects on observed variants were performed using the online software SIFT (http://sift.bii.a-star.edu.sg) and SNAP2 (https://www.rostlab.org/services/snap/). Integration sites were analysed using the QuickMapper online program provided by the Gene Therapy Safety group at gtsg.org.

2.6 Imaging

2.6.1 Immunofluorescence

Expression of MLV- Gag in tested cell lines was evaluated with the use of a primary MLV Gag antibody (rabbit) and a secondary anti-rabbit goat IgG labelled with Alexa-488 (ab100970, Abcam, UK). The cells were grown in a glass-free Nunc Lab-Tek Chamber Slide system, on a Permanox slide (VWR, UK). Negative controls were included omitting the addition of the primary antibody (negative primary control) as well as both antibodies (negative antibody control). All antibodies were diluted in a solution containing 0.2% (w/v) fish skin gelatin (Sigma G-7765), 0.02% (w/v) saponin, 0.02% (w/v) NaN3 in PBS. Cells were then fixed, using 3% paraformaldehyde in PBS solution, permeabilised and quenched using 50mM NH4Cl,

0.2% (w/v) saponin in PBS. After mounting with Mowiol (BioRad, UK), the chambers were visualised in a Nikon fluorescence microscope or a Leica confocal microscope as indicated.

Human stem cells were labelled additionally with a primary Anti-Integrin A6 antibody (17-0495 eBioscience, UK) that is permits visualisation of cells that have maintained their undifferentiated stem state (expressed in rat). Human corneal stem cells were also labeled against anti-cytokeratin K3/K12 (rabbit, 2Q1040, Abcam,UK), a specific antibody for staining differentiated keratin. The use of eucaryotic Phalloidin (eFluor-570 -41-6559, eBioscience) was also evaluated in human and mouse cells for cytoskeleton stains.

2.6.2 Immunofluorescence using the indicator fluorescent DERSE cell line

iGFP-DERSE cells (named after the acronym Detectors of Exogenous Retroviral Sequence Elements) are subclones of LnCap prostate cancer cells that have been stably transfected with a construct plasmid, capable of acting as a reporter gene, i.e. turning cells green fluorescent in case of a replication-competent retrovirus present.

Cells are grown to confluence (>70%) on 9mm glass coverslips and are then used in the indicator cell assay. In detail, supernatant containing putative replication-competent retrovirus is used to super-infect the LnCap DERSE cells at variable titres and cells are incubated (37^{o}C 5% CO_2) for at least 15 days. Cell-bearing coverslips are then mounted for microscopy (according to the Immunofluorescence protocol- section 2.12.1) at 1,2,4,6,10 and 15 days post infection. eGFP expression is then quantified as fluorescent cells/coverslip.

2.6.3 Transmission electron microscopy

Cells were grown on 9mm glass coverslips and fixed with 2.0 % w/v paraformaldehyde and 1% w/v glutaraldehyde in DMEM growth medium. They were stained as per a standard TEM protocol in Public Health England, Colindale. Post-fixation was carried out with 1% aqueous osmium tetroxide (OsO4) and potassium ferricyanide (1.5% w/v), then washing with sodium cacodylate and the cells were stored. The cell bearing coverslips were then embedded in Epon resin. Ultra- thin (grey < 60 nm) sections were cut with a diamond knife, and stained with 1% Uranyl Acetate before imaging with a Hitachi H7000 transmission electron microscope (Hitachi, Tokyo, Japan).

Figure 2.6 The DERSE reporter-cell circular plasmid

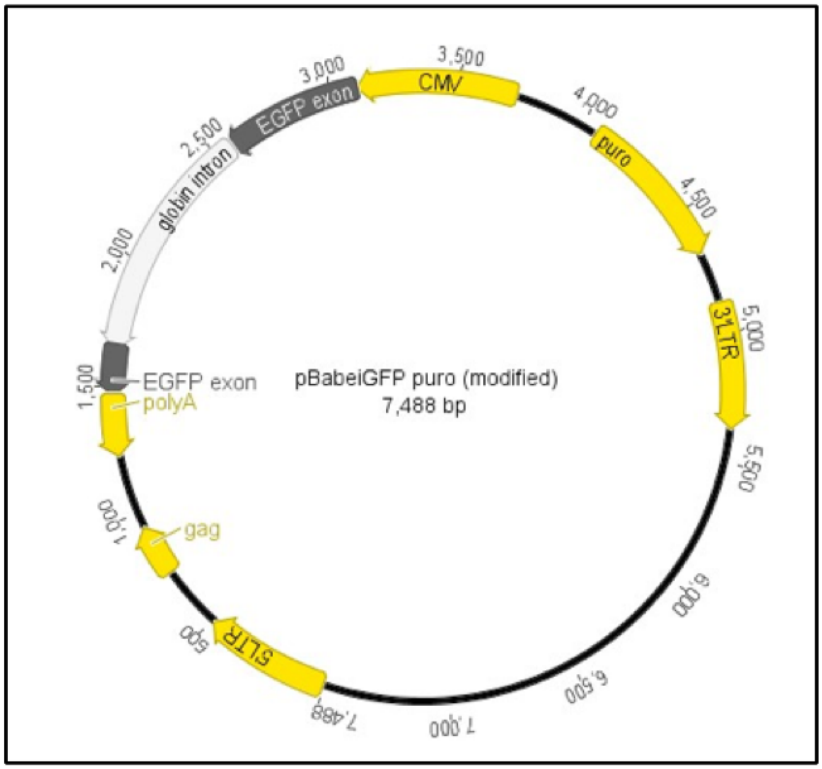


Figure 2.6 The DERSE reporter-cell circular plasmid: An eGFP sequence and CMV promoter is placed in reverse direction with respect to flanking MoMLV-LTRs. The plasmid is used to stably transfect LnCap cells that are selectable via Puromycin resistance (Puro). Within the eGFP gene an intron (human globulin gene) has been additionally placed to inactivate it at the DNA level. Splice donor and acceptor sequences are placed on the edges of the globulin gene, allowing it to be removed with splicing when the construct is transcribed as an RNA intermediate. The transfected iGFP reporter cells produce genomic RNAs with spliced, minus-strand eGFP-CMV genes that accumulate in the cytoplasm, but cannot code for functional protein (i.e. do not fluoresce).

Chapter 3
Investigation of MLV virus activation in murine cell lines

3.1 Introduction

Early in 1960, it was apparent that non-defective latent endogenous retroviruses of the MLV class in mice could be activated i.e. detected in RNA form, either spontaneously or post cell treatment with several inducing agents. Such an activation was observed initially after whole radiation treatment of virus-free mice and soon after, the chemical compound Bromodeoxyuridine (BrDU) was demonstrated to activate MLV endogenous retroviruses in mouse fibroblasts (Margalith et al, 1975). The reaction to inducing stimulants however, was found to vary, as expected, according to the genetic reservoir of MLV proviruses in the mouse strains tested; fluctuating to dose alterations or drug combinations. After an extensive review of the endogenous retrovirus activation literature in the 1960s, it is obvious that a wide array of agents are able to act as stimulators of a previously inactive endogenous retrovirus albeit, viral induction may be restricted to a very small percentage of treated cells (as low as 0.01-0.03%) and indeed be dose-dependent (for a comprehensive literature review of the two main chemical stimulation compounds please see Appendixes 3 and 4).

In general, MLV inducing agents are chemicals that are able to abolish endogenous virus latency via the following actions:

a) Demethylation or inhibition of the methylation of DNA over chromosomal areas that have been previously silenced. Such agents include iododeoxyuridine, bromodeoxyuridine and 5-azacytidine; as well as the more recent histone deacetylase inhibitors (HDAC).

b) Direct action on molecular enhancer sequences present in retroviruses. Such agents include glucocorticoids, such as dexamethasone that acts on the glycocorticoid responsive element (GRE) of retroviruses.

c) Other less-described activation functions: These activation agents include chemicals such as tuftsin, kentsin, hormones, cordycepin, lipopolysaccharide (LPS), Mitomycin C, tunicamycin and irradiation or cell starvation.

Several protocols have been employed by nationally competent authorities such as the FDA for the detection of occult endogenous retroviruses in cells (Khan et al, 2001); these protocols are mainly designed to inform over the safety of primary cells used in the production of biologicals or other controlled material, such as vaccines (Switzer et al, 2011). This investigation in MEF feeder cells however has not been rigorous, nor included in standard good manufacturing protocols as of yet; especially in the case of mouse to human cell contact experimentation as is the case of human stem cell co-culture prior transplantation (Chapter 7).

I primarily aimed to investigate the ability of several MEF cell lines, used in human stem cell culture and transplantation, to produce activated and replication competent endogenous retroviruses after the application of a range of chemical treatments. I here first describe the action of the two most commonly used inducing agents in MEF cell culture, with the gold reference protocol for this experiment being the one described by the FDA and, present with my results.

3.1.1 5-Iododeoxyuridine (5-IDU)

5-Iododeoxyuridine (5-IDU) is a halogenated nucleotide that has traditionally been used to facilitate the investigation of DNA structure using X-ray crystallography and protein-DNA complexes via UV-crosslinking. It causes the demethylation of silenced gene regions in genomic DNA via direct incorporation into DNA (Willis et al, 1993).

The demethylating action of 5-IDU has been applied in endogenous retrovirus research as it can efficiently stimulate the multifold-release of several inactive endogenous retroviruses in mouse cells (Khan et al, 2001). After a brief literature review that aimed to identify the optimal dosage, duration of treatment and mouse strains most susceptible to the chemical activation effect of 5-IDU in vitro, I evaluated twenty-seven studies that have documented sufficient endogenous retrovirus release in NIH-3T3, Balb-c, C57BL6 or AKR mouse fibroblasts via 5-IDU treatment alone or on combination with other chemicals (Appendix 3). Of these, the

majority of researchers treated cells for 24 hours and a virus peak was observed via RT detection either in the 2nd or 3rd day post treatment. The dosage varied from 20-100μg/ml however one study alone, described a high cell toxicity observed with concentrations above 30μg/ml (Khan et al, 2001).

5-IDU is not soluble in water and is reconstituted in NH4OH (at concentration 1N), a vehicle that has been shown to also have a toxic effect on cell culture. Apart from toxicity however, another side-effect is concerns over the inhibition of viral infectivity, as iododeoxyuridine may introduce errors in DNA replication thus inducing the release of defective virions (Patch et al, 1981). Washing of the cells post treatment as well as the inclusion of a NH4OH only control was therefore necessary in all 5-IDU experimentation for the control of these effects.

3.1.2 5- Azacytidine- C (5-AZAC)

5-azacytidine-C is a cytidine analog that acts as a DNA methyltransferase inhibitor, thus blocking the methylation of genomic DNA in areas that become silenced (Hagemann et al,2011). It is one of many azanucleosides that have been studied for their action in silent gene activation in normal and cancerous tissues. After its cellular uptake, the molecule is activated and converted into an active nucleotide that causes DNA methyltransferase inhibition; the active molecule is 5-aza-2-deoxycytidine-5-triphosphate. This is incorporated into DNA and substitutes for cytosine. DNA methyltrasferase normally establishes a covalent bond with cytosine rings in the position of carbon-6, however the presence of a nitrogen molecule present in azacytidine in the neighbouring place of cytosine's carbon-5, blocks this parallel binding, and thus the action of the methyltransferase is inhibited.

Inhibition of DNA methylation has often been proposed as a mechanism for the re-activation of latent viruses (Chandler et al, 1988), and the methylation of transcriptionally active viral regions such as the HIV-1 LTR has been considered an important step in the establishment of latency (Ishida et al, 2006). Recently it has

been shown that the action of DNA methyltransferase inhibitors is not random (Flotho et al, 2009), but rather specific, non-random and reproducible and maybe directed at promoter-rich regions rather than gene-body CpG-islands (regions rich in —C—phosphate—G— i.e. CpG). In mouse cells, 5-Azacytidine C has also been historically associated with the induction of type A intracisternal retroviral particles (IAP) in various cell types (Lasneret et al, 1983), a type of retrotransposon (Reuss et al, 1996). Although present in many copies (>1000) in mouse genomes, IAP particles are not considered infectious, albeit they might be associated with increased oncogenicity (Maksakova et al, 2006).

The role of 5-Azacytidine-C in the activation of fully replication-competent retroviruses in mouse embryonic fibroblasts was assessed prior to experimentation. I performed a review of the literature up to date to identify the optimal dose and conditions required for the induction of endogenous retroviruses from various mouse cells by cytidine analogs and in particular 5-Azacytidine-C (Appendix Table 4). Six (6) studies evaluated 5-Azacytidine-C alone or in combination with other chemicals for mouse fibroblast cells. Viral induction in reviewed studies was detected at days 1, 2 and 3, while the effect was shown to be dose-dependent and effective in >90% of treated cells particularly in AKR cells.

3.2 Pre-assessment of experimental conditions

3.2.1 Growth curve characteristics of MEF cells

Mouse embryonic fibroblast (MEF) cells J2-3T3 (Altrika, UK), MEF-CF1 (ATCC SCRC-1040), BL6 (ATCC SCRC-1008), Balb-K (ATCC CL-163.3), NIH-3T3 (ECAAC No 930 61524) primarily stored in liquid nitrogen were thawed as described in Methods 2.1, and set up to grow in 6- multi-well plates in 1ml of DMEM growth medium supplemented with 10% FBS that was previously heat-inactivated at 56^{o} for 30 min and 1% pen/strep. Cells were maintained in a humidified incubator at 37^{o} with 5% CO_2.

The population doubling time of the NIH-3T3 cell lines in culture, was determined in T25 flasks over 72 hours over three different plating densities ($1x10^6$ cells, $0.75x10^6$ cells and $0.5x10^6$ cells), using trypan blue exclusion staining for proliferation counts, in a standard haemocytometer plate (mean of four measurements as cells/mm^3). The remaining cell lines were plated in a standard $0.5x10^6$ density. The calculation of lag time and doubling time was performed to determine optimal chemical stimulation time points, shown to occur when treatment is applied at approximately the start of the log phase of cell growth (Results shown in Table 3.2.1) For the prototype cell line, the lag phase is concluded in the first 20hours for all initial seeding densities; while at 20 hours the exponential log phase starts. The calculation of the population doubling time (PDT) is performed during the log phase and it equals 14.8, 15.6 and 18.6 hours for this cell line (depending on the initial seeding density), and is on average 16.4h±2h. The measurements show that the CF1-MEF cells are slower at recovery from the seeding phase with a lag phase of >24h (varying from 24 to 36h depending on batch of CF1 cells tested) and a population doubling time (PDT) of 18.5h, while the rest of the murine feeder cell lines have comparable characteristics with the prototype NIH-3T3 line.

Table 3.2.1 The growth curve characteristics of the cell lines used in the induction experiments.

Cell line	initial density (cells/well)	lag	mean doubling time
NIH-3T3	5E+05; 7.5E+05;1E+06	20h	16.4h
CF1-MEF	5E+05	>24h	18.5h
J2-3T3	5E+05	20h	18.5h
BL6-MEF	5E+05	24h	20.5h

Figure 3.2.1 Growth curve determination for the prototype NIH-3T3 cell line

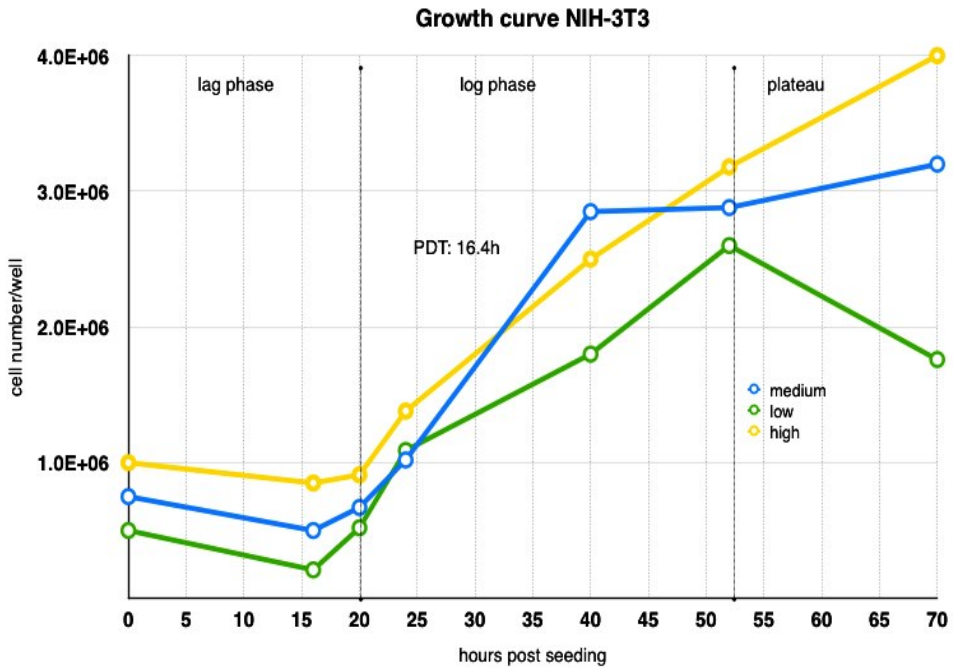


Figure 3.2.1 Growth curve determination for the prototype NIH-3T3 cell line

Growths of three initial seeding densities i.e. $1x10^6$ cells in yellow, $0.75x10^6$ cells in blue and $0.5x10^6$ cells in green, were measured at 16h, 20h, 24h, 40h and 48h post seeding. The characteristics of the cell line's lag phase, log phase and plateau or confluence phase, are depicted in the growth curve graph..

3.2.2 Drug toxicity and effect on cell proliferation of MEF cells

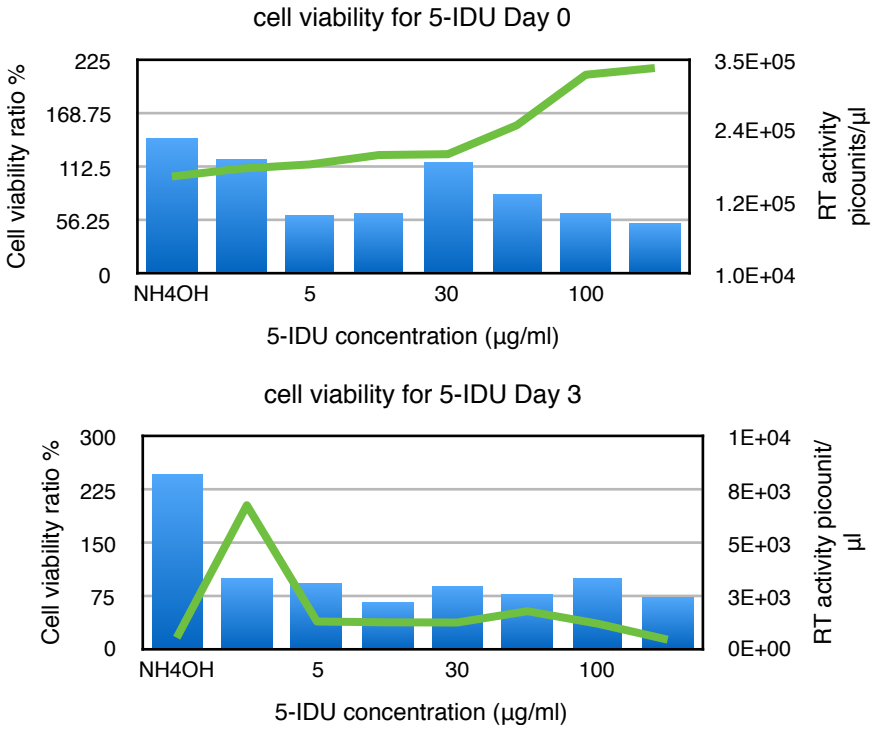
Drug toxicity and cell proliferation over three days after an initial 24h chemical treatment, were determined in 24 well plates using trypan blue exclusion in a standard haemocytometer plate. The inducing drugs tested were reconstituted and stored as per the manufacturer's instructions. Drugs and conditions used were 5-Iododeoxyuridine (Sigma, I7125), 5-Azacytidine C (Sigma, A2385), LPS1 (E.Coli 0111:B) (Sigma), LPS2 (E.Coli 0111:A) (Sigma), Mitomycin C (Roche, MN), Hydroxyurea (Sigma) and Dexamethasone (Sigma) and combinations of the above. The line tested was initially NIH-3T3 cells and subsequently CF1-MEF, BL6-MEF, and J2-3T3 cells in specified drugs and dosage range.

Cells were grown for 24 hours in 24-well plates (seeded initially with $5x10^4$ cells) and then treated with varying concentrations of inducing drugs for 24 hours, after which the drugs were removed, cells were washed 3x with PBS and the viable cell number was determined as described above. The day immediately post 24-hour treatment, after drug removal was designated Day 0 -post removal of drug. Cell proliferation was evaluated until Day 3 after drug removal, and compared against untreated controls and drug solvent vehicle controls for the estimation of toxicity. In particular, the cell viability ratio % was calculated by dividing the number of surviving cells in treated wells to the number of surviving cells in untreated wells (%).

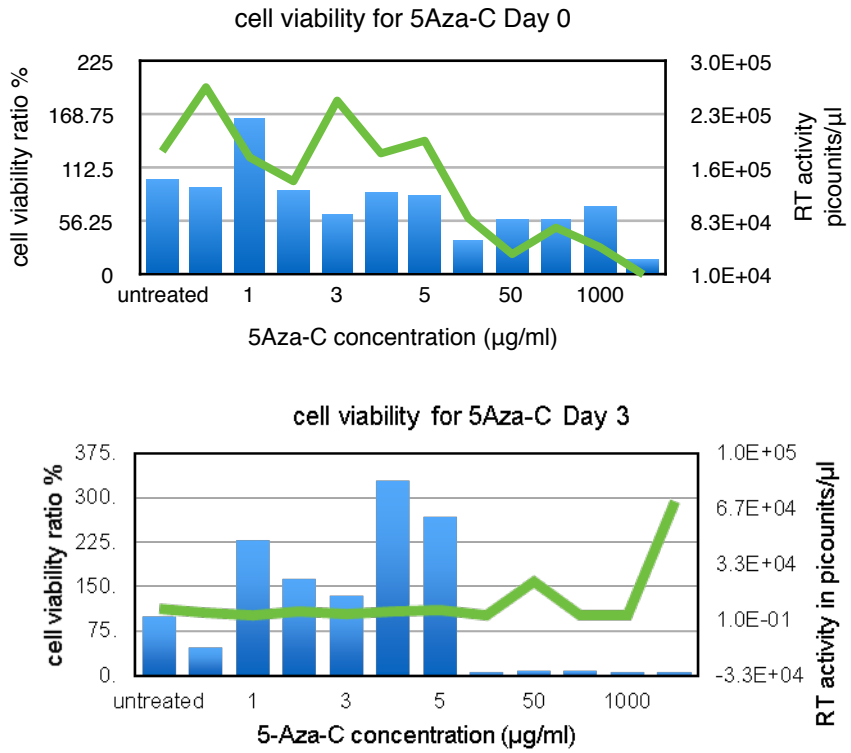
Drug concentrations were eligible for consideration in the formal induction experiment provided that cell proliferation exceeded 50% in day 0 post removal of the drug and 75% in day 3. Samples were also screened by PERT for the detection of RT activity for day 0 and day 3 which was the indicative recovery evaluation period post treatment (see following section).

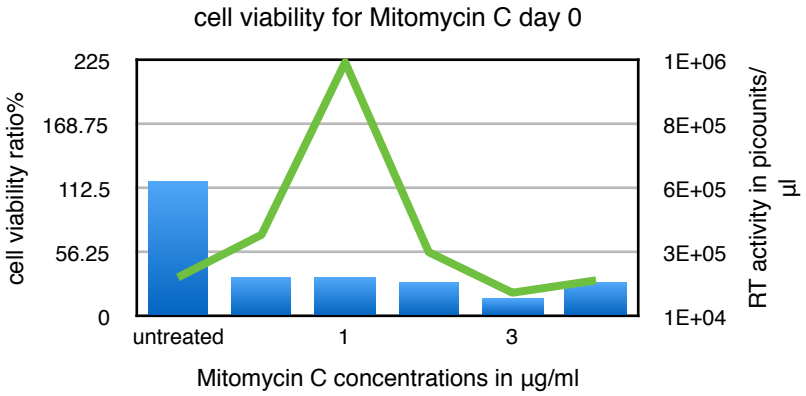
Figure 3.2.2 Cell viability assays on NIH-3T3 cells with varying concentrations of trial inducing agents

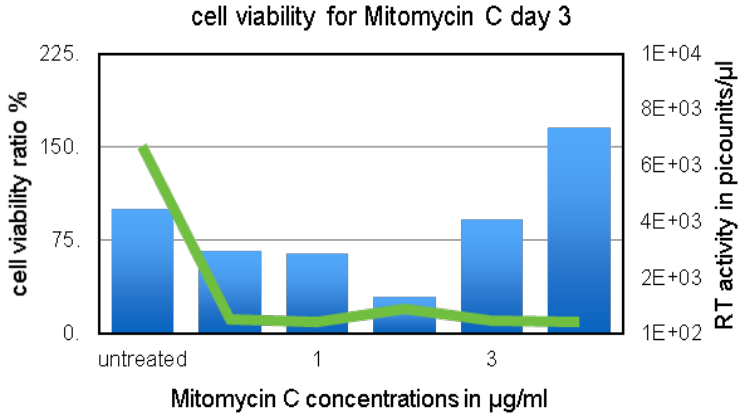
A.



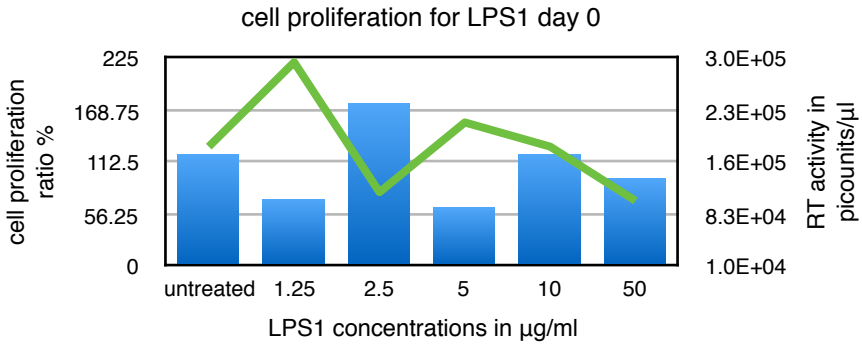
B.

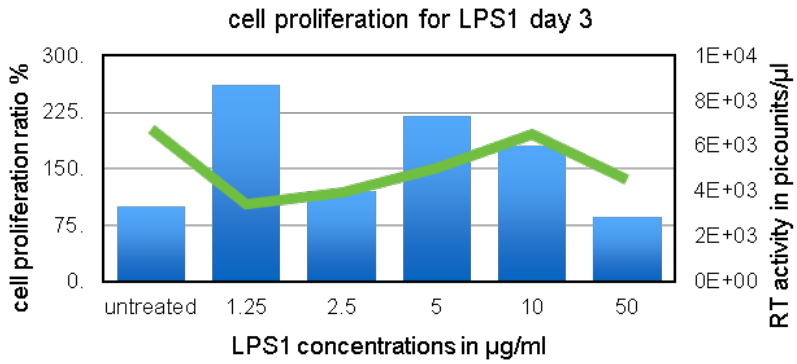



cell viability for Mitomycin C day 0
cell viability ratio%
225
168.75
112.5
56.25
0
1E+06
8E+05
6E+05
3E+05
1E+04
RT activity in picounits/ µl
untreated
1
3
Mitomycin C concentrations in µg/ml


cell viability for Mitomycin C day 3
cell viability ratio %
225.
150.
75.
0.
1E+04
8E+03
6E+03
4E+03
2E+03
1E+02
RT activity in picounits/µl
untreated
1
3
Mitomycin C concentrations in µg/ml

C.


cell proliferation for LPS1 day 0
cell proliferation ratio %
225
168.75
112.5
56.25
0
3.0E+05
2.3E+05
1.6E+05
8.3E+04
1.0E+04
RT activity in picounits/µl
untreated
1.25
2.5
5
10
50
LPS1 concentrations in µg/ml


cell proliferation for LPS1 day 3
cell proliferation ratio %
300.
225.
150.
75.
0.
1E+04
8E+03
6E+03
4E+03
2E+03
0E+00
RT activity in picounits/µl
untreated
1.25
2.5
5
10
50
LPS1 concentrations in µg/ml

D.

E.

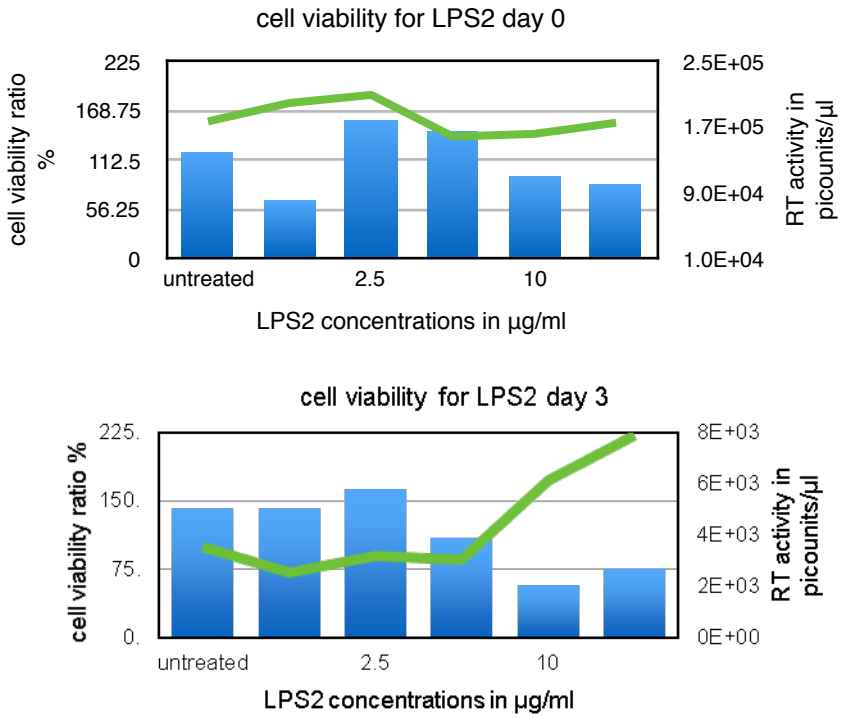


Figure 3.2.2 Cell viability assays on NIH-3T3 cells with varying concentrations of trial inducing agents

(Experiments A-E) The cells were counted on day 0 and day 3 post 24-hour treatment with inducing concentrations of chemicals. Live/dead cell counts were performed via trypan blue exclusion and cell viability ratios were calculated as surviving cells indexed against the surviving control untreated cells (blue bars). The green line depicts the PERT assay measurement results of the overall RT activity in cell supernatants in picounits/µl. 5-IDU: 5-iododeoxyuridine, 5-AzaC: 5-azacytidine C, Mitomycin C, LPS1: lipopolysaccharide from E.Coli O111B4, LPS2: lipopolysaccharide from E.Coli O127B8.

It is unlikely, from previous publications, that mouse fibroblasts would need as high as 200µg/ml of 5-IDU alone for the induction of retroviruses and these combinations were shown to be highly toxic and non-viable in my conditions (data not shown). A similar lack of cell proliferation was observed for all tested concentrations of dexamethasone, hydroxyurea and concentrations of Mitomycin C above 5µg/ml, and the particular chemical variables were excluded from further analysis.

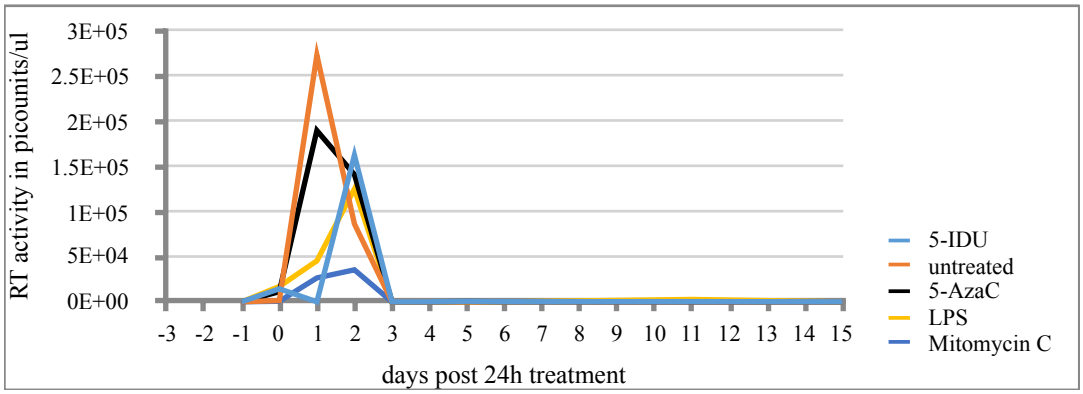
3.2.3 Screening PERT assays for candidate chemical stimulants

PERT assays were performed on cell supernatant from the cell culture wells corresponding to cell viability testing of varying chemical concentrations, on day 0 post 24hour treatment and day 3 post treatment. MEF cell lines (CF1-MEF, BL6-MEF and NIH-3T3) were grown in 6-well plates for 24h and were subsequently treated with various concentrations of chemicals, as previously discussed (Section 3.2.2). Controls consisted of untreated cells while a NH4OH vehicle control corresponding to 30µg/ml of 5-IDU was also included. Cells were washed with PBS after removal of drug post 24h (day 0) and supernatant harvested until day 21 for PERT analysis of RT activity.

Having already determined the maximum concentration of chemicals which did not display significant toxicity (ie 5-IDU 30µg/ml, 5-AzaC 1.25µg/ml, Mitomycin C 4µg/ml and lipopolysaccharide (LPS) 10µg/ml as per Figure 3.2.2) I performed PERT measurements for the test cell lines over a longer period of time i.e 21 days (Figure 3.2.3). The ability to induce endogenous MLV viruses, as demonstrated via the detected rise in RT activity is achieved in our experimental protocol for NIH-3T3 cells and 5-AzaC or LPS (unique peaks of RT activity at day 1 post treatment) and for CF1-MEF cells and 5-AzaC (peaks at day 0-1-3). NIH-3T3 cells were lost post day 15 for unknown reasons. Another concern identified is that although cells were washed at day 0 post treatment for 3 times to reduce the likelihood of the drug remnant passing onto the PERT assay, this event could not be excluded.

Figure 3.2.3 Evaluation of selected chemicals on MEF cell lines for the induction of RT activity over 21 days

A.



B.

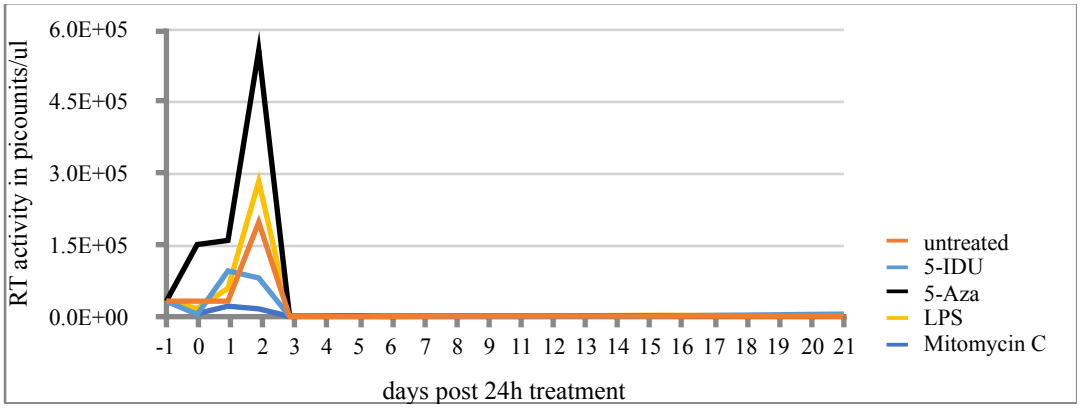


Figure 3.2.3 Evaluation of selected chemicals on MEF cell lines for the induction of RT activity over 21 days.

A: NIH-3T3 and

B: CF1-MEF

Cells were treated with selected quantities of inducing drugs (5-IDU 30µg/ml, 5-AzaC 1.25µg/ml, Mitomycin C 4µg/ml and lipopolysaccharide (LPS1 from E.Coli O111B4) 10µg/ml) and sample supernatants were assessed for RT activity in singleton measurements. Controls included growth medium alone and NH4OH treated cells (not shown). Several peaks in RT activity are distinguished indicative of the possibility of a one stage viral induction in the experimental protocol.

Cells died after day 15 in A.

3.3 Induction experiment

For the formal induction experiment, the following cell lines were tested, in triplicate:

1/ **CF1-MEF** cells (stem cell experimentation in wide use);

2/ **J2-3T3** (currently used in human keratinocyte stem cell expansion and provided via the UK Stem cell Bank) and

3/ the positive induction control cell line **Balb-K** (ATCC).

Although the J2-3T3 cell line was not included in the preliminary test drug induction due to timing constraints, the characteristics of the cells were considered equivalent to the previously available NIH-3T3 cells as per common origin, and were further tested under similar conditions.

Cells were grown for 24 hours in 24-well plates (seeded with $5x10^4$ cells) and then treated with inducing drugs for 24 hours, after which the drugs were removed, 1ml of supernatant was harvested and stored in 2ml screw-cap Starstaed tubes. Immediately post treatment (day 0) cells were washed 3x with PBS. The cultures were replenished with fresh medium and re-incubated at 37^o 5% CO_2. Sampling of tissue supernatant continued for 21 days post-treatment, at 3-day intervals after day 4.

Iododeoxyuridine (5-IDU) was applied, after reconstitution in NH4OH, in three different concentrations for tested MEF cell lines (20-30-50μg/ml) and at 20μg/ml for Balb-K cells. 5-AzaC was applied in three different concentrations for MEF cell lines (1.25-2-5 μg/ml) and at 2μg/ml for Balb-K cells. Experimental variance in the actual chemical concentrations may possibly be confounding reproducibility due to intra-laboratory variation, inadvertent dilution errors or technical cross-variation and last, possibility of minimal vehicle-remnant passing onto the experiment post-diligent wash.

3.3.1 Assessment of the induction of endogenous MLVs in the established cell line Balb-K

To assess the efficiency of my cell culture conditions to induce endogenous retroviruses from mouse cells, confluent Balb-K cells growing in 6-well plates in 3ml of growth medium were first treated with 30µg/ml 5-IDU and 2µg/ml of 5-AzaC for 24 hours. The induction agents were removed on day 0, cells were washed and growth medium replaced. Supernatant samples were collected from day 0 (day of the removal of the drug) until day 19, first every 24 hours and following day 4, every 3 days, in the following manner. One (1) ml of supernatant was removed replenished with 1 ml of growth medium. Non-induced controls were untreated Balb-K cells. Cell supernatants were tested by PERT and RT-PCR in triplicate.

Cell culture supernatant was analysed for RT and amplified for MLV *gag* sequences according to Methods (Section 2.5.3 Page 66). Standard curves for quantification were constructed for PERT with Mo-MLV RT and for RT-PCR with freshly extracted 22rv1 cell supernatant extract from growing 22rv1 cells containing XMRV.

The viral induction in Balb-K cells from 5-IDU evolves indeed rapidly with RT activity increasing on day 0 to reach $6x10^7$ picounits/µl one day earlier than an increase is detected by MLV *gag* RT-PCR, as seen in figure 3.3.1A (PERT: above and MLV *gag* RT-PCR: below).

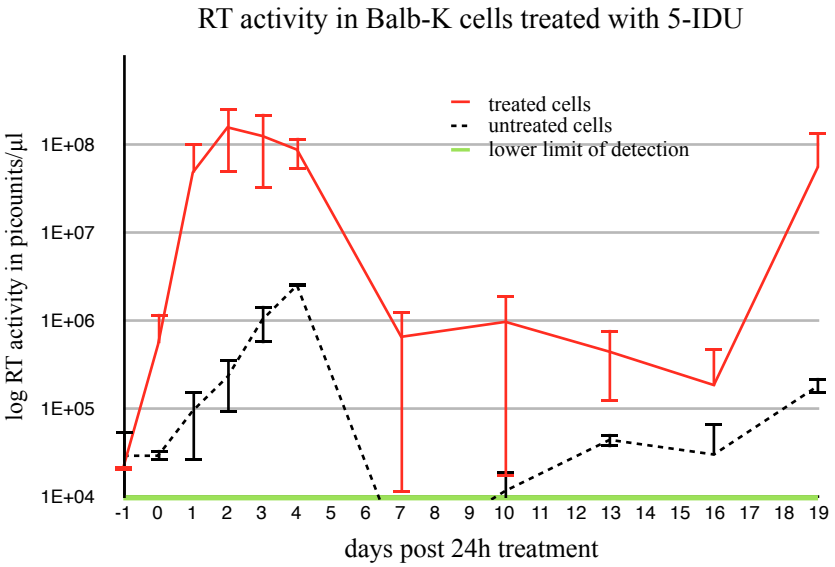
From Figure 3.3.1A the Balb-K cells treated with 5-IDU exhibit a large surge in RT activity on day 1 post induction, peaking on day 2 at $2x10^8$ picounits/µl and remaining equally high for days 3 and 4. From day 7 cells are showing less significant levels of enzymatic activity while a second peak is seen on day 19 (10^8 picounits/µl). Untreated negative Balb-K cell controls, (black dotted line) are consistently lower than the significant level of detection of our PERT assay (10^4 picounits/µl) (data not shown). This upsurge is comparable to the values observed in Khan et al, 2001 (peak of 10^8 picounits/µl day 1 post induction).

In 5-IDU treated Balb-K cells, the peak in MLV *gag* RNA starts on day 3 with an excess of $8x10^4$ viral copies of MLV *gag* RNA and on day 7 reaches $2x10^5$ copies/µl, a titre that falls to zero in the following days. The results show an increase in viral copies starting from day 2 and peaking in day 7. In day 7, lack of a second peak in RT activity does not support a second distinct virus release. A lower increase is seen in the RT activity detected in 5-AzaC treatment (peak of 10^6 picounits/µl on day 7 compared to $1.5x10^7$ picounits/µl measured in Khan, 2001 on day 1) (Figure 3.3.1.B: top). This marginal increase of RT activity over untreated cells is supported by a distinct peak in circulating MLV *gag* RNA from day 1, day 3, 4 and 7 (maximum of $9x10^3$ RNA copies/µl) suggesting that an inducible MLV virus may indeed be present.

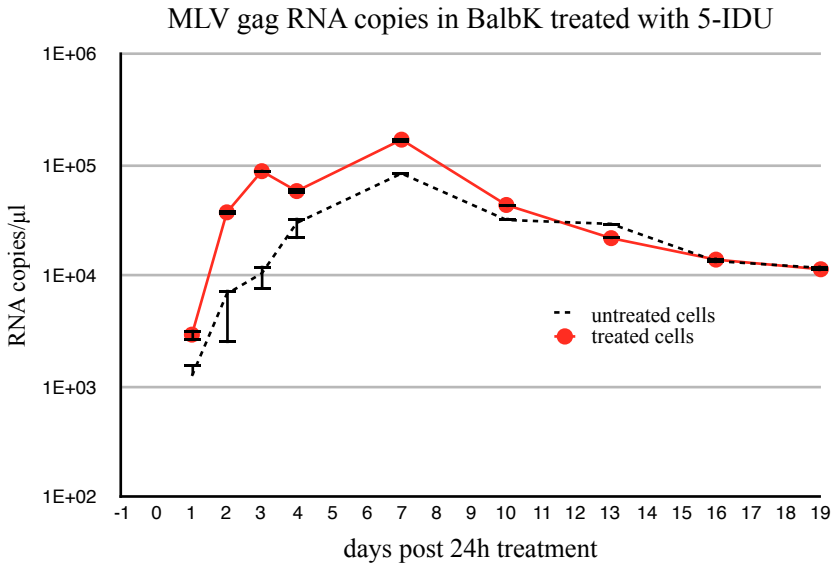
The RT activity level of $5x10^4$ picounits/µl in untreated cells on average per total untreated days is indeed confirmed by Balb-K cells in both 5-IDU and 5-AzaC experiments. This level of RT activity is slightly above (0.5log10) our PERT detection limit of 10^4 picounits/µl.

Figure 3.3.1: RT activity and MLV viral load in Balb-K cells

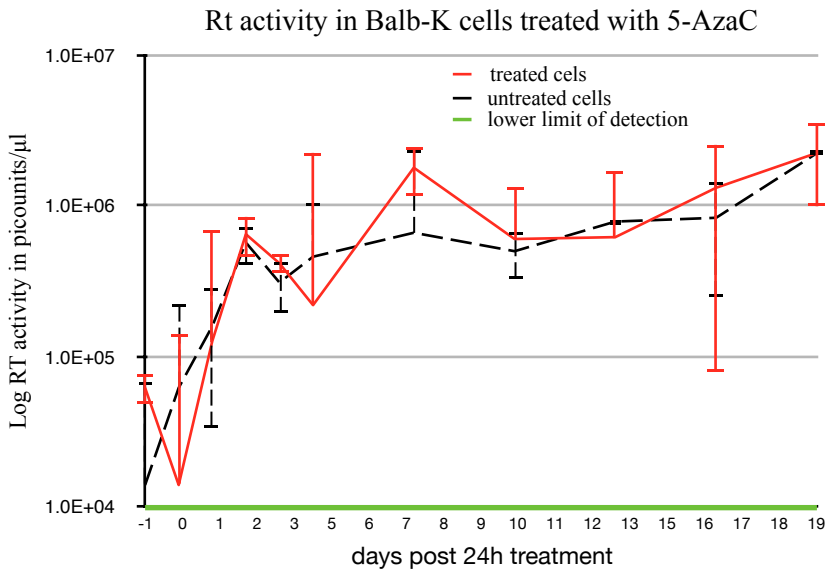
A.



B.



C.


Rt activity in Balb-K cells treated with 5-AzaC
Log RT activity in picounits/µl
1.0E+07
1.0E+06
1.0E+05
1.0E+04
treated cels
untreated cells
lower limit of detection
-1 0 1 2 3 5 6 7 8 9 10 11 12 13 15 16 17 18 19
days post 24h treatment

D.

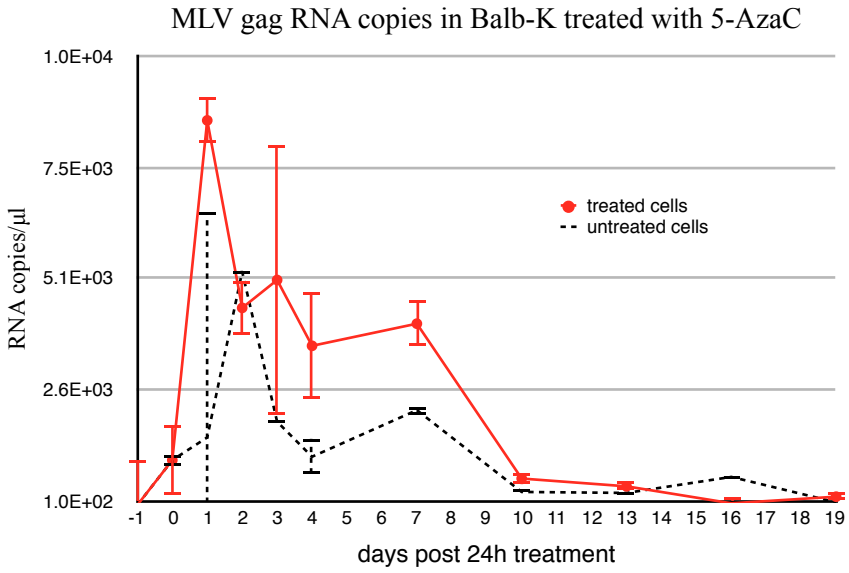

MLV gag RNA copies in Balb-K treated with 5-AzaC
RNA copies/µl
1.0E+04
7.5E+03
5.1E+03
2.6E+03
1.0E+02
treated cells
untreated cells
-1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
days post 24h treatment

Figure 3.3.1 RT activity and MLV viral load in Balb-K cells

A. **RT activity in Balb-K cells treated with 30µg/ml 5-IDU.** A significant increase is observed for treated cells (red line) compared to untreated cells (dotted black line). The peak is highest on day 2 post treatment at a value of $2x10^8$ picounits/µl. The results are plotted over the minimum of 10^4 picounits/ µl , which the lower detection limit of the PERT assay (green line). Error bars represent standard deviation of triplicate cell culture well. A standard curve of Mo-MLV RT enzyme from 10^{-2} to 10^{-8} IU/µl was used for PERT quantification (slope :3.92, R^2: 0.998) (not shown).

B. **MLV *gag* RNA copies in Balb-K cells treated with 30µg/ml 5-IDU.** An MLV RNA signal is observed in the treated cells in days 3 and 7 post treatment (red line) versus untreated cells (black dotted line). All time points are the mean of three replicate measurements. Error bars represent standard deviation of triplicate cell culture well. A standard quantification curve was obtained through serial ten-fold dilutions of a same-run freshly extracted extract from 22rv1 cells containing XMRV (1 million to ten copies/ PCR reaction) (slope: 3.22, R^2:0.99) (not shown). RNA differential copies are calculated using the formula $2^{-\Delta Ct}$ in RT TaqMan PCR.

C. **RT activity in Balb-K cells treated with 2µg/ml 5-AzaC.** A very small increase is observed for treated cells (red line) and is comparable to the increase in untreated cells (dotted black line). The results are plotted over the minimum of 10^4 picounits/µl , which the lower detection limit of the PERT assay (green line). Error bars represent standard deviation of triplicate cell culture well. A standard curve of Mo-MLV RT enzyme from 10^{-2} to 10^{-8} IU/µl was used for PERT quantification (slope :3.95, R^2: 0.989) (not shown).

D. **MLV *gag* RNA copies in Balb-K cells treated with 2µg/ml 5-AzaC.** An MLV RNA signal is almost undetectable at mean 8000 copies/µl in the treated cells in day 1 post treatment (red line) versus untreated cells (black dotted line). All time points are the mean of three replicate measurements. Error bars represent standard deviation of triplicate cell culture well. A standard quantification curve was obtained through serial ten-fold dilutions of a same-run freshly extracted extract from 22rv1 cells containing XMRV (1 million to ten copies/ PCR reaction) (slope: 3.22, R^2:0.99) (not shown). RNA differential copies are calculated using the formula $2^{-\Delta Ct}$ in RT TaqMan PCR.

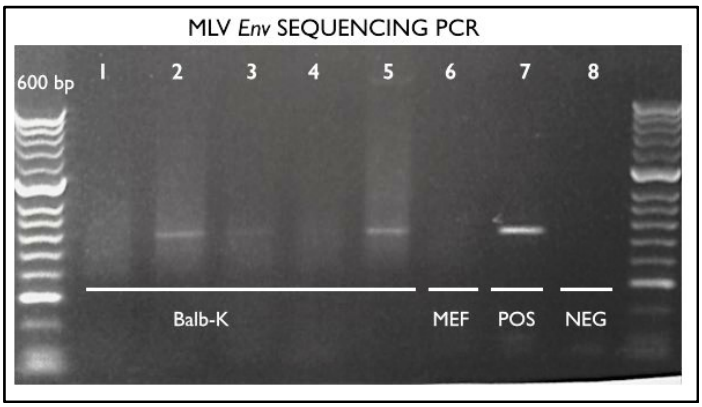
3.3.2 Confirmation of molecular sequences involved in the Balb-K induction

Viral nucleic acids (NA) were extracted from induced Balb-K supernatants and amplified with primers spanning the *gag* and *env* region, as well as with Kirsten-Murine Sarcoma Virus p21 gene primers. (See Appendix 1 for list of primers). In three successive amplification runs, MLV *gag* nested PCR produced an amplicon of the expected product size from Balb-K cells in 5-IDU or 5AZA-C conditions. Sequencing however proved impossible using MLV primers, but p21 primers directed at the predominant murine sarcoma virus present in Balb-K cells were successful.

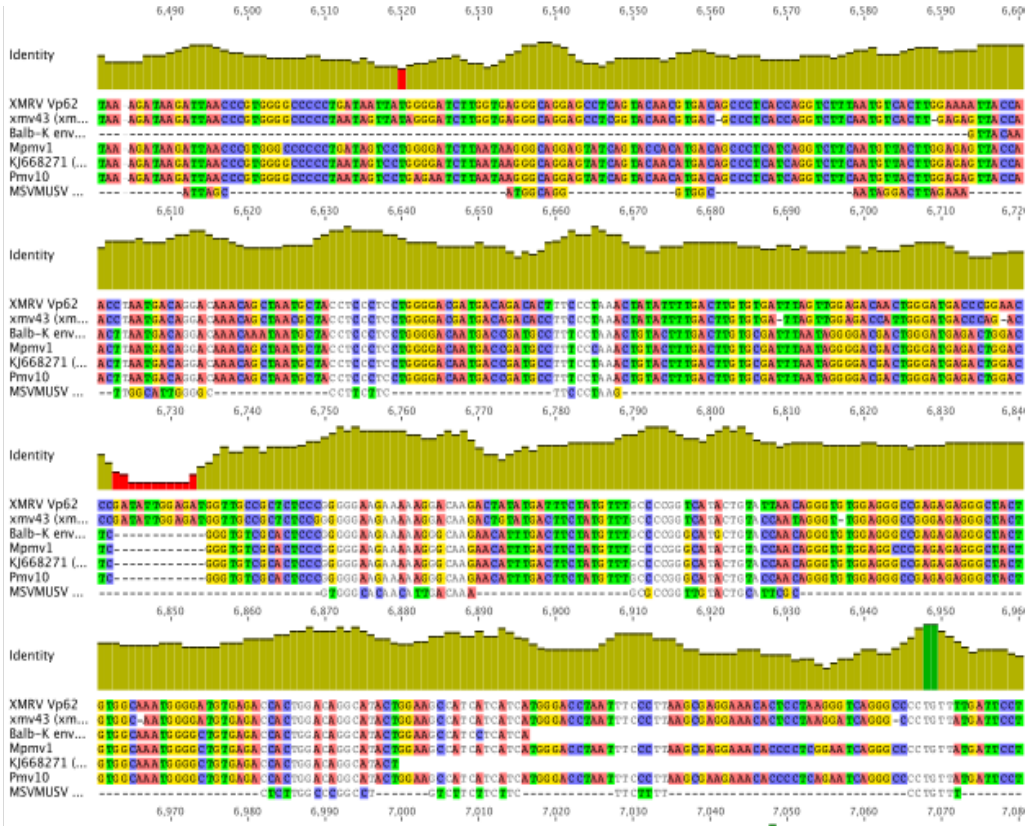
In envelope gene PCR (Figure 3.3.2A) amplicons from all Balb-K supernatants (5-AzaC treated and 5-IDU treated) were produced. These were successfully sequenced and gave an MLV-like envelope amplicon of the expected size at 600 bp (envelope primer target range approx. 600-700bp; range to allow for isolate variation). The supernatant extracts tested repeatedly gave identical sequencing results in the *envelope* region that aligned with a polytropic endogenous MLV (pMLV10) (Jern et al, 2007).

Additionally, using the primers described in Methods for XMRV/p-MLV g*ag* sequences, designed by Tuke et al, 2011 we were able to amplify with Real time-PCR a significant number of x/p-MLV single strand (ss) *gag* RNA copies in the Balb-K cells tested (up to $5x10^3$ ss RNA copies/µl). Samples used were the peaks from Figure 3.3.1.A and Figure 3.3.1.C at days 3-7-10-13 and days 2-3-4 and 19; further strengthening our result that this is a polytropic or xenotropic virus in Balb-K cells consistent with previous reports. The xMLV *gag* amplicons however failed to amplify at direct sequencing (data not shown). These primers being designed for Taq-RT-PCR indiscriminately failed to produce sequencable products from all samples tested.

Figure 3.3.2 Envelope sequencing PCRs of BALB-K induced supernatants



A.



B.

Figure 3.3.2 Envelope sequencing PCR of Balb-K induced supernatant

A. MLV *env* gene sequencing for Balb-K induced supernatants: Lanes 1 Balb-K in 5-AzaC day 3, Lane 2: Balb-K untreated day 3, lanes 3: Balb-K in 5-AzaC day 13, lane 4: Balb-K untreated day 13, 5 Balb-K in 5-IDU. Three Balb-K samples (treated with 5-AzaC day 13, treated with 5-IDU day 2 and untreated day 3) produced bands that were sequenced via Sanger direct sequencing, with amplicons of the expected product size 600bp.

B. Alignment of Balb-K derived *env* sequences (Lanes 2,3 and 5 in Figure 3.3.2A) produced envelope sequences that aligned 100% amongst them and variably with MLV envelope sequences. This alignment's consensus Balb-K *env* is found more closely related to the polytropic endogenous MLVs (pMLV1) compared on individual base pair polymorphisms than xmv-43, that is considered traditionally to be the inducible virus in Balb-K cells. The particular insertion observed for xenotropic envelopes XMRV and xmv-43 in position 6845 to 6854 is characteristically missing. The alignments were performed in Geneious.

.

3.3.3 Assessment of the induction of endogenous MLVs in J2-3T3 cells

J2-MEF cells were evaluated for the release of RT activity in lysed cell supernatant after treatment with three varying concentrations of 5-AzaC (1.25-2.5-5 µg/ml) and 5-IDU (20-30-50 µg/ml). Although the RT activity in J2 cells +5-IDU, J2 cells +5-AzaC and untreated, were above the limit of detection in the PERT assay (10^4 picounits/µl), the increase was less than 1log10 in maximum peaks in day 2 and 3 (with 20µg/ml 5-IDU and 50µg/ml 5-IDU) or day 7 (with 5-AzaC 5µg/ml). A peak of uncertain significance in day 7 post 5-AzaC treatment is followed by a similar surge from untreated cells and is not evaluated further (Figure 3.3.3.1C).

Similarly, the RT-qPCR on J2 MEF cell supernatant extracts did not produce significant results for the number of copies of MLV *gag* RNA detected and quantified in RTase (+) conditions. The quantification resulted in an average of 500 RNA copies of MLV *gag* on days 1-13 in J2+5-IDU.

Induced and un-induced extracts from time points of higher RNA content or RT activity were investigated with *gag* and *env* sequencing PCR (Figure 3.3.3.2). The J2 cell supernatant from day 7 untreated, day 10 untreated, day 3 treated with 5-IDU 30mg/µl and day 10 treated with 5-AzaC 5mg/µl was amplified with *gag* sequencing primers. In MLV *gag*, J2 related endogenous MLV amplicons aligned closer to polytropic endogenous MLV (specifically pmv-9 and preXMRV2; e-value:1e-10^5), an endogenous defective polytropic MLV that is not replication competent, than XMRV or ecotropic MLVs, but this region in *gag* has no significant discriminatory value for phylogenetic analysis.

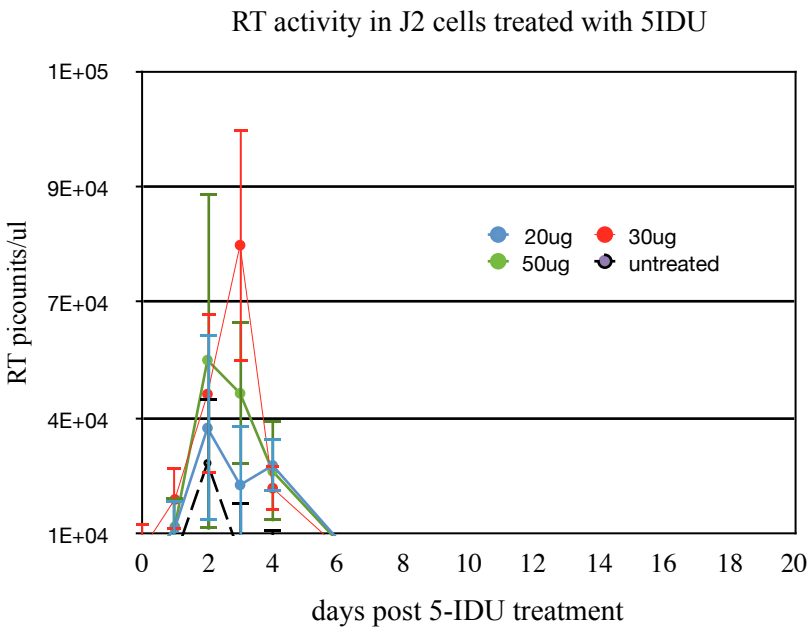
The *env* sequencing results for RNA and DNA substrates from the same chosen J2 induced cells (ie with and without the addition of RTase) show significant consensus amongst all sequences from J2 cell supernatant with and without RT, with 99% pair wise identity (Figure 3.3.3.2). The consensus J2 *env* sequence aligns with various endogenous polytropic envelopes of the Mus Musculus genome, as well as with pre-XMRV2, and does not align with xenotropic sequences like XMV-43 or pre-

XMRV1. The annotated base pair positions in the J2 *env* consensus (Figure 3.3.3.2B) are determinants of a modified polytropic nature (Evans et al, 2009); in these positions the J2 envelope sequence bears 2 our of 3 SNPs similar to modified polytropic viruses and 3 same SNPs similar to pre-XMRV2, a defective non-replication-competent polytropic MLV.

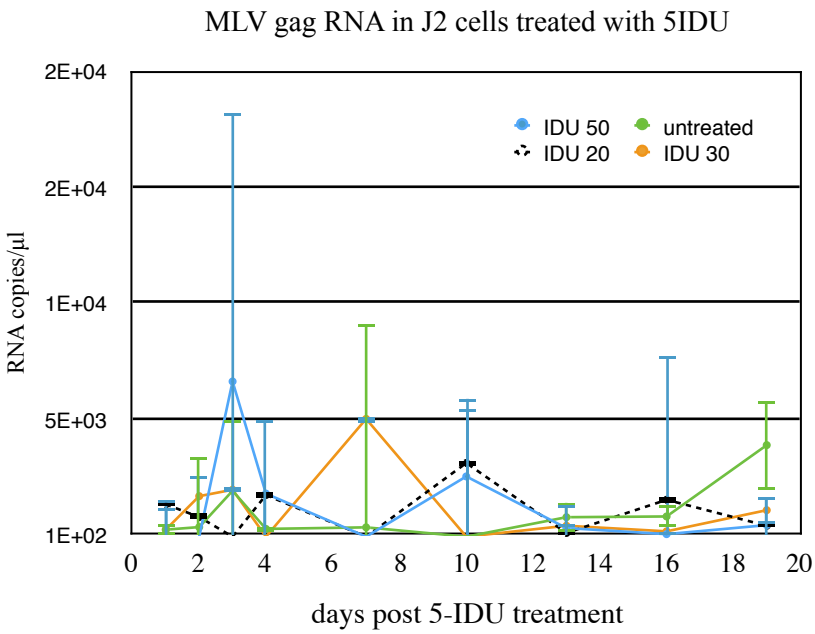
However, this finding coupled with the lack of inducible RT activity observed during 19 days in stimulated J2 cells suggests that although expression is seen, an activation of a replication-competent polytropic virus in J2 cells is unlikely. Indeed, polytropic viruses are also unable to infect cells on their own. These findings are discussed further in the discussion section.

Figure 3.3.3.1 RT activity and MLV viral load in J2 cells treated with 5-IDU

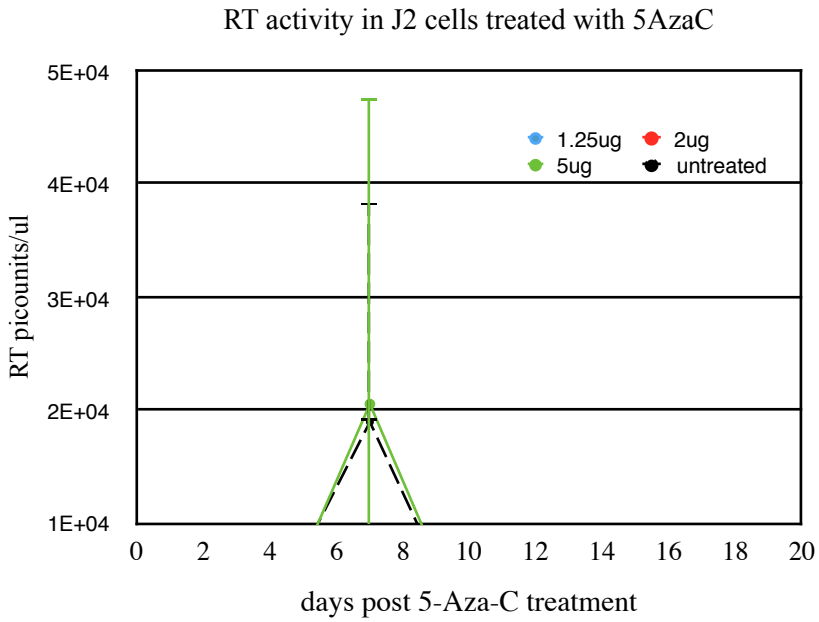
A.

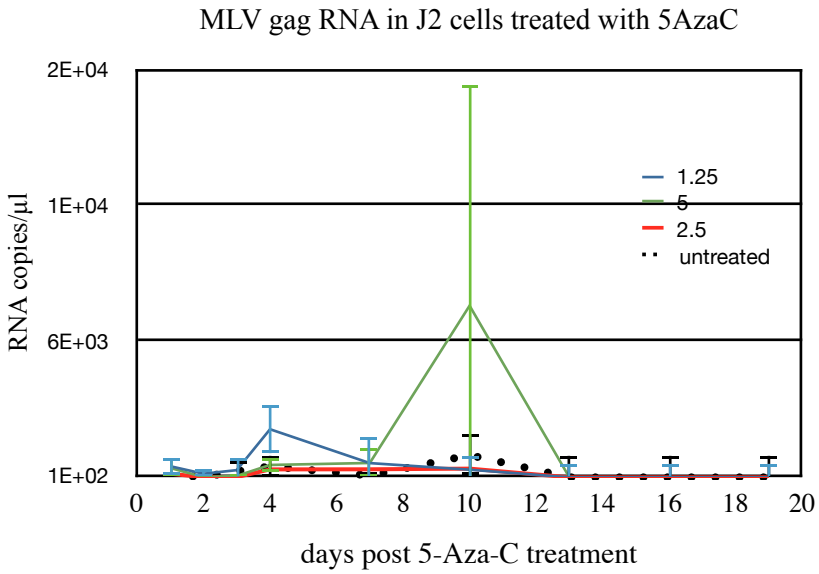


B.



C.


RT activity in J2 cells treated with 5AzaC
RT picounits/ul
5E+04
4E+04
3E+04
2E+04
1E+04
1.25ug
2ug
5ug
untreated
0
2
4
6
8
10
12
14
16
18
20
days post 5-Aza-C treatment


MLV gag RNA in J2 cells treated with 5AzaC
RNA copies/µl
2E+04
1E+04
6E+03
1E+02
1.25
5
2.5
untreated
0
2
4
6
8
10
12
14
16
18
20
days post 5-Aza-C treatment

D.

Figure 3.3.3.1 RT activity and MLV viral load in J2 cells treated with 5-IDU

A. RT activity in J2 cells treated with 5-IDU. Three different 5-IDU concentrations were evaluated for their action on J2 3T3 cells in triplicate (red, green and blue) versus untreated cell supernatant associated RT activity (black dotted line). Error bars represent standard deviation. The limit of detection of the PERT assay is $\mathbf{10^4}$ picounits/µl at baseline.

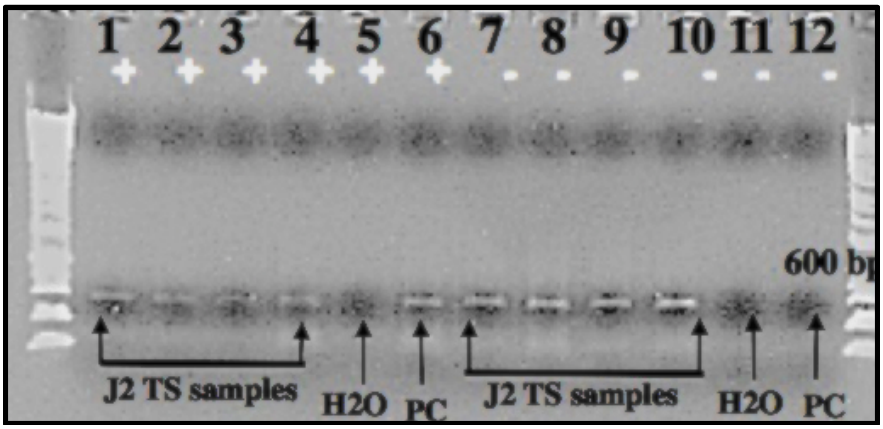
B. MLV *gag* RT-PCR in J2 cells treated with 5-IDU (20-30-50µg/ml). The extracts were assessed by RT (+) real time PCR and RT (-) real time PCR to allow for the calculation of differential RNA content. Error bars represent standard deviation. The standard curve for quantification was obtained with serial dilutions of 22rv1 cell supernatant extract (1million to 10 copies/PCR reaction) (Slope: -3.1 R2:0.99) (not shown).

C. RT activity in J2 cells after treatment with 5-AzaC. Three different 5-AzaC concentrations were evaluated for their action on J2 3T3 cells in triplicate in red, blue and green versus untreated cells (black dotted line) for 20 days post treatment. Error bars represent standard deviation. The limit of detection of the PERT assay is 10^4 picounits/µl at baseline.

D. MLV *gag* RT-PCR on J2 cells treated with 5-AzaC (1.25-2.5-5 µg/µl). The extracts were assessed by RT (+) real time PCR and RT (-) real time PCR to allow for the calculation of differential RNA content via ΔΔCT. The standard curve for quantification was obtained by serial dilutions of 22rv1 cell supernatant extract (1 million to 10 copies/PCR reaction) (slope: -3.57, R2: 0.983) (not shown).

Figure 3.3.3.2 MLV *env* gene sequencing for J2 cell supernatant

A.



B.

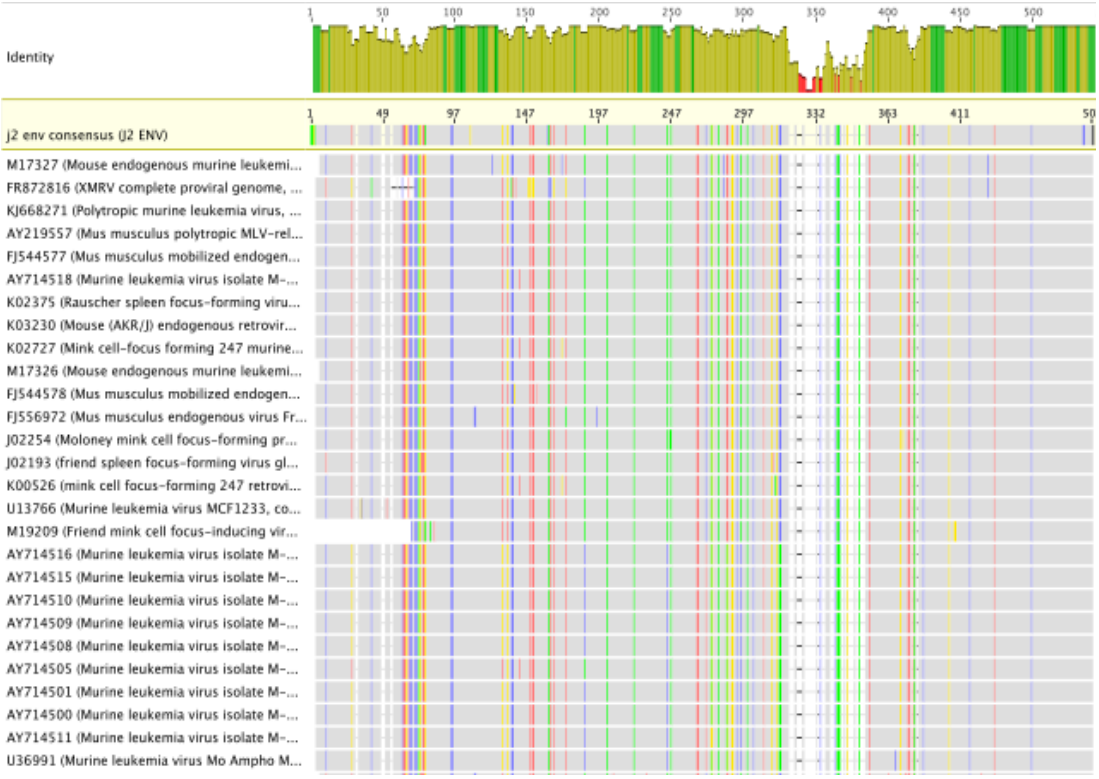


Figure 3.3.3.2 MLV-*env* gene sequencing PCR of J2 cell supernatant

J2 cell supernatants (TS: Tissue supernatant) from MLV *gag* RNA peaks (Figure 3.3.3.1A bottom and 3.3.3.1B bottom), were amplified with *env* sequencing primers with (+) and without (-) reverse transcriptase synthesis.

A. Amplification results in RT+ and RT- PCR: Lanes 1-4 and 7-10 are J2 (Lane 1: 5-AzaC 1.25µg day 2, Lane 2: 5-AzaC 1.25µg day 10, Lane 3 5-IDU 30µg day 1, Lane 4 5-IDU 30µg day 10), lane 6 and 12 is the positive control (PC) 22rv1 RNA (XMRV) and lanes 5 and 11 are water negative controls. The J2 cell supernatant extracts have a high proportion of gDNA that is amplified in the RT (-) wells with a solid band; the positive control is negative for RT (-) amplification as the 22rv1 cells contain a significant copy number of XMRV RNA (>10^8 copies/µl).

B. Alignment of MLV *env* derived sequences for J2 cell supernatant. The sequencing results showed significant alignment of J2 cell supernatant from all days, with 99% pair wise identity that allowed for the production of a consensus sequence. The consensus J2 *env* sequence aligns with various endogenous polytropic envelopes of the Mus Musculus genome, as well as with pre-XMRV2, but it does not align with xenotropic sequences such as XMV-43 or pre-XMRV1. Several positions in the J2 envelope consensus sequence are found to be determinants of a modified polytropic nature (see text).

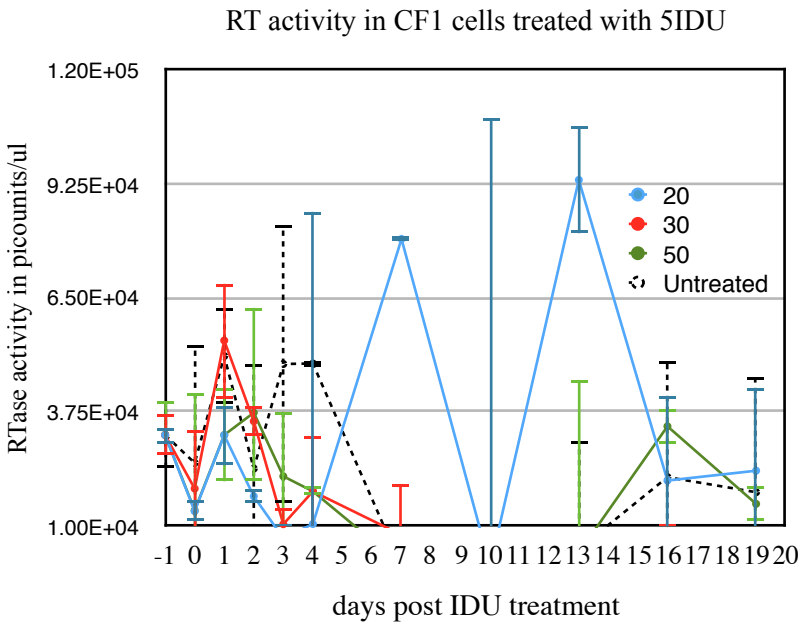
3.3.4 Assessment of the induction of endogenous MLVs in CF1 cells

CF1-MEF cells were batched and archived for consistency as different batches may have different endogenous retrovirus activation (Takeuchi et al, 2008a). The cells were treated in three different concentrations of 5-AzaC and 5-IDU. RT activity was measured in lysed induced supernatant samples collected for 20 days post 24-hour treatment and stored in -80^{o}C. The results of the PERT assays are summarised in Figure 3.3.4A and 3.3.4B (top). A dual excursion above borderline in RT activity is observed in CF1 MEF cells treated with 20μg/ml of 5-IDU versus no RT activity in the two remaining 5-IDU concentrations (30μg/ml and 50μg/ml) and untreated cells (Figure 3.3.4A). The peaks in day 7 and 13 (10^5 picounits/μl) are not comparable to the maximum peak of RT activity on day 3 for the positive control 5-IDU treated Balb-K cells, which is above 10^7 picounits/μl. An equally small increase in RT activity is observed in 5-AzaC treated CF1 cell supernatant, on the 4th day post treatment with all three concentrations (1.25-2.5 and 5μg/ml) compared to untreated cells at $3x10^5$ picounits/μl. Untreated cells have consistently low RT activity at the limit of detection (10^4 picounits/μl).

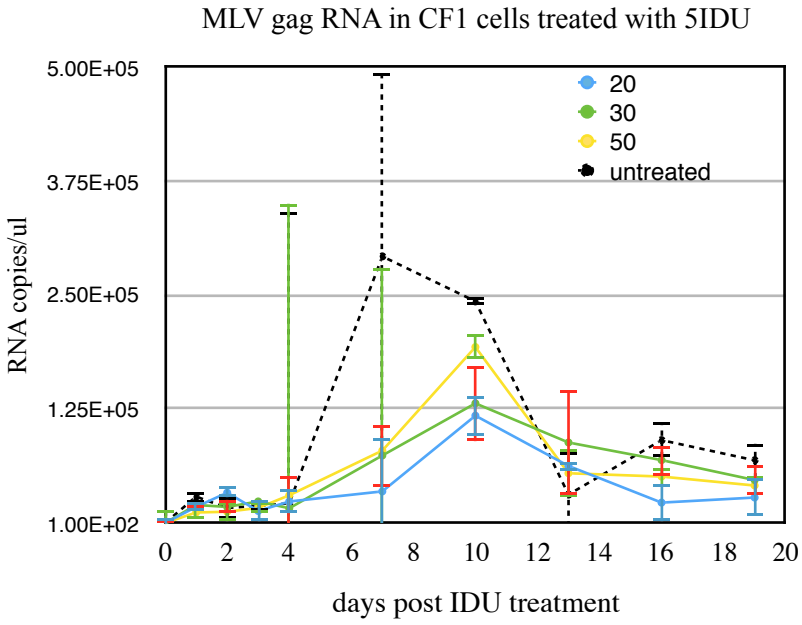
Standardised 96-well plate Taq-qPCRs with (RT+) and without (RT-) were performed on extracts from the automated extraction of viral nucleic acids for 5-AzaC and 5-IDU treated CF1 MEF supernatant and untreated controls. (Figure 3.3.4B and 3.3.4D bottom). MLV *gag* RNA is present in all treated and untreated CF1 supernatant extracts at titres above $5x10^3$ RNA copies/μl in all conditions and reaches peaks of 10^5 to $5x10^5$ in untreated cells. In detail, all RT positive extracts exhibit an earlier cycle threshold value on average 6.7 cycles in 5-AzaC treated and untreated CF1 cells and 8.8 cycles in 5-IDU treated and untreated CF1 cells, establishing the perception of a constant CF1 related MLV *gag* RNA element or elements present. This result is indicative of one or multiple expressed CF1- related MLV elements, that however are not associated with high retroviral RT activity as demonstrated in the PERT results (Figure 3.3.4A and C). Also, the relative peaks of RT activity for days 7 and 13 for 5-IDU and day 7 for 5-AzaC are not associated with a higher MLV *gag* RNA titre, as quantified in the Taq- qPCR for the particular post-treatment day extracts.

Figure 3.3.4 Induction of CF1 cells with 5-IDU and 5-AzaC: results of PERT and RT-PCR

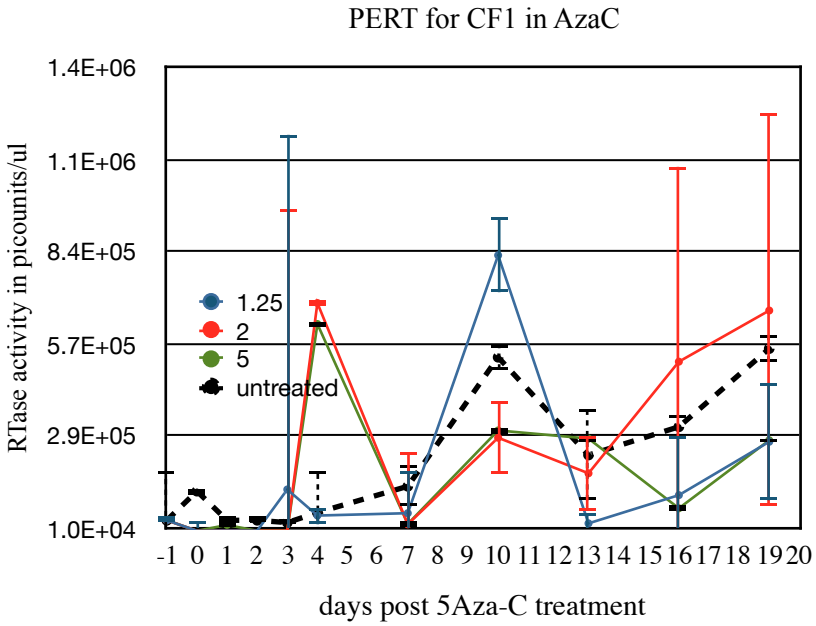
A.



B.



C.


PERT for CF1 in AzaC
RTase activity in picounits/ul
1.4E+06
1.1E+06
8.4E+05
5.7E+05
2.9E+05
1.0E+04
1.25
2
5
untreated
-1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
days post 5Aza-C treatment

D.

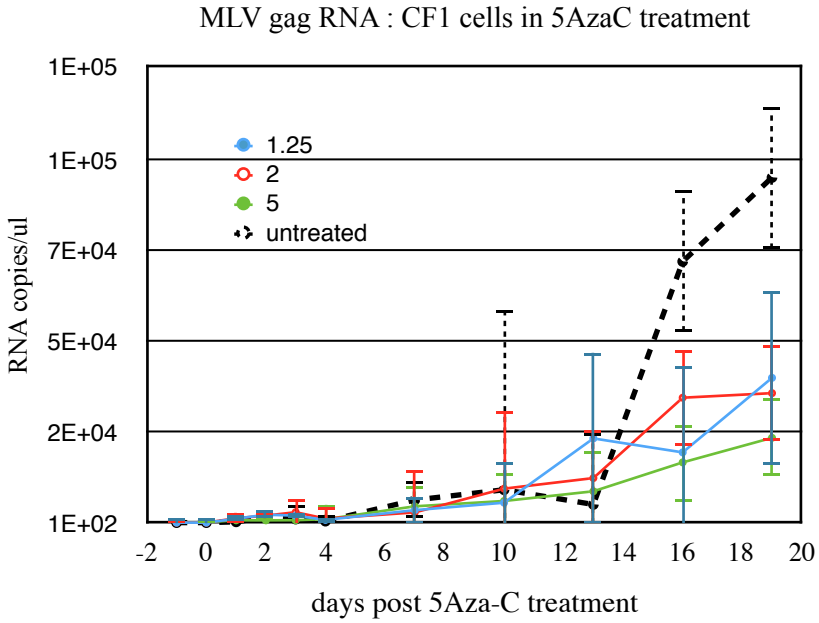

MLV gag RNA : CF1 cells in 5AzaC treatment
RNA copies/ul
1E+05
1E+05
7E+04
5E+04
2E+04
1E+02
1.25
2
5
untreated
-2 0 2 4 6 8 10 12 14 16 18 20
days post 5Aza-C treatment

Figure 3.3.4 Induction of CF1 cells with 5-IDU and 5-AzaC: results of PERT and RT-PCR

A. RT activity measured from CF-1 cell supernatant after treatment with 5-iododeoxyuridine (5-IDU) for 24h. Supernatant was collected from day 0 (removal of drug) until 20 days post-treatment. Three different 5-IDU concentrations were evaluated for their action on CF-1 MEF cells seeded in 6-well plates (10^5 cells /well) in triplicate (red, blue and green lines). Untreated cell supernatant RT activity is shown in a black dotted line. Error bars represent standard deviations. The assay's limit of detection is 10^4 picounits/µl shown as baseline.

B. RT (+) real time PCR and RT (-) real time PCR for MLV *gag* RNA quantification in CF1-5-IDU treated cells (ΔΔCT). The MLV mixed RNA+DNA signal detected was on average: 8.8 CTs higher (3.13log2 higher) than the DNA signal for days 1-19, while the documented average cycle threshold for RNA is equivalent to 2500 MLV *gag* ssRNA copies/µl. The 22rv1 cell extract RT (+) standard curve has a slope of -3.24 and R2 of 0.939 (not shown).

C. RT activity measured from CF-1 cell supernatant after treatment with 5 azacytidine C (5-AzaC) for 24h. Supernatant was collected from day 0 (removal of drug) until 20 days post- treatment. Three different 5-IDU concentrations were evaluated for their action on CF-1 MEF cells seeded in 6-well plates (10^5 cells /well), in triplicate (red, blue and green lines). Untreated cell supernatant is shown with the black dotted line. Error bars represent standard deviation. The assay's limit of detection is 10^4 picounits/µl shown as baseline.

D. RT (+) real time PCR and RT (-) real time PCR for MLV *gag* RNA quantification in CF1-5-AzaC treated cells (ΔΔCT). The MLV RNA+DNA viral nucleic acid signal was on average: 6.771 CTs higher than DNA (2.7log2 higher) while the average cycle threshold is equivalent to 200 MLV *gag* ssRNA copies/µl. The 22rv1 cell extract RT (+) standard curve has a slope of -2.99 and R2 0.959 (not shown).

3.4 Discussion

MEF feeder layers have been known to bear endogenous MLV viruses in numerous copies and on occasions activated proviruses have been discovered in random batches of such cell lines (Takeuchi et al, 2008a). Screening for the potential activation of endogenous retroviruses via chemical stimulation is not an official routine test performed before the safe release of MEF cells in clinical or other research use in the UK. However, as novel viruses appear such as XMRV, we undertook this risk assessment exercise for two of the most widely used mouse cell lines (J2 and CF1). These were chosen from our preliminary test experiments among three different MEF cell lines selected for their ability to produce initial RT activity in preliminary test conditions.

The report of toxicity in the 30µg/ml 5-IDU induction concentration (Khan et al, 2001) was not observed in my conditions. When looking at the cell viability ratio in the NH4OH treated cells, a control that was set up to correspond to the amount present in the mean 30µg/ml 5-IDU concentration (3.8x10-3N), it is also obvious that there is no attributable effect on cell viability for concentrations tested.

Two separate preliminary assessments of the RT activity in tested cells were made to inform the formal induction experiment. The results of the RT activity detection are shown in Figure 3.2.2, parallel to cell viability assessment. Subsequently, I tested the target cell lines under investigation (CF1-MEF, BL6-MEF and NIH-3T3) with a subset of non-toxic chemical concentrations for an initial assessment of PERT activity over 21 days (in singleton wells). Several peaks in RT activity were distinguished indicative of the possibility of a viral induction in the experimental protocol.

The positive induction control cell line, Balb-K exhibited comparable results in our experimental conditions to the published literature, with exceptionally high RT activity measured in induced versus untreated cells, particularly for 5-IDU chemical stimulation, where I measure $6x10^7$ picounits /µl on day 0 and $2x10^8$ picounits/µl in

day 1 post induction in treated cells. This upsurge is comparable to the values detected in Khan et al, 2001 that measures a peak of 10^8 picounits/µl day 1 post induction with 5-IDU. The remaining measurements are also comparable, such as the second peak on day 19 with 5-IDU, validating our experimental protocol according to the established positive induction control.

The induced Balb-K supernatant extracts tested, moreover repeatedly produced sequencing results in the envelope region (using conserved envelope primers) that aligned unanimously with a polytropic endogenous MLV (pMLV10) (Jern et al, 2007).

Treated and untreated J2 cells show a very low level of RT activity and viral RNA that becomes undetectable post day 4 and 10 respectively. Their content in endogenous MLVs however suggests that a polytropic endogenous retrovirus is present and polytropic sequence determinants are demonstrated for *env*. Interestingly, the content of the J2 cells in endogenous MLVs has post these experiments, now less relative interest with regards to the safety of stem cell co-cultivation practice, as even if polytropic viruses are induced they cannot become competent being that they require the presence of an exogenous ecotropic or xenotropic retroviral infection to propagate (Evans et al, 2009) (Young et al, 2012).

Treated and untreated CF1 cells demonstrate higher RT activity throughout two separate measurements conducted in triplicate, that is on average $7X10^4$ picounits/µl for a duration of 20 days. The higher RT activity in days post day 7 in both conditions, may be consistent with a concentration of viral particles, as growth medium is also replenished every 3 days after day 4, compared to daily in the first 4 days of measurement. However, the highly conserved MLV *gag* RNA calculation via real time RT-PCR across the samples-days demonstrates a detectable RNA signal throughout the 20 days that is on average 30,000 ss RNA copies/µl in untreated CF1 cells. This is likely indicative of a constituent MLV activated virus in CF1 cells prior to and after chemical treatment.

The finding that untreated CF1 cells have on average the same MLV *gag* RNA content in both experiments is further proof of our hypothesis of a presence of a unique MLV virus in CF-1 cells. This is further pursued in the following Chapter. The difference in copy numbers between 5-AzaC and 5-IDU treated supernatant (8.000 copies/µl versus 10.000 copies/µl) appears to be either due to a secondary MLV activation (or IAP activation as sometimes seen with 5-Aza-C historically) or is random; as RT activity has peaked conversely slightly higher in 5-AzaC versus 5-IDU treated samples. The specific sequence of the MLV-RNA in CF1 cells is the subject of our next chapter, where we attempt to identify and characterise it.

Treated CF1 cells additionally demonstrated a slight rise in RT activity for 20µg/ml 5-IDU (for days 7 and 10) and in all 5-AzaC concentrations (for days 4 and 10) and a gradual RT upsurge with all concentrations of 5-AzaC after day 7, which were 1log10 higher for 5-AzaC compared to 5-IDU treatment (average $3x10^5$ picounits/µl compared to $3x10^4$ picounits/µl) across the observation days. The question of whether this observation is indicative of an additional virus activation has to be considered on the basis of 'peaks' being significantly low both to the induction control and very close to our PERT assay's limit of detection. This is further visited via particular time point direct sequencing of extracted nucleic acid extracts in the following chapter.

We attempted to PCR amplify the genomic content of CF1 cells using the current primers described in this Chapter and with a wide array of new LTR primers, hoping to fully characterise the content of the CF1 cells in endogenous proviral copies. However, after 4 months of experiments we failed to obtain PCR products in our primer walking strategy. (data not shown).

During the experiments described herein, I have hence recorded a constituent MLV virus expression in CF1-MEF cells, using highly conserved MLV *gag* primers that I describe and investigate extensively in the following Chapters, also taking advantage of the increased detection sensitivity of next generation sequencing. The lack of RT activity associated with this MLV *gag* RNA signal from CF1 cells is a factor in the consideration of the CF1 MEF virus being possibly replication defective.

Chapter 4
Molecular and virological characterisation of the CF1 cell line-related MLV

4.1 Introduction

In this chapter, the MLV-RNA element that was detected in CF1-MEF cells in Chapter 3 via RT-PCR for MLV *gag* is further investigated using further molecular PCR assays and sequencing, including novel full RNA-virus next generation sequencing. An ecotropic MLV found abundant in CF1 cells is aligned and compared to the previously published sequences from CF-1 mice and similar strains.

4.2 RESULTS

4.2.1 Identification of CF1 related MLV-RNA using MLV *gag* (Nested) PCR

Using a nested-PCR approach to amplify MLV *gag* (product size for first round 334 bp and for second round 110 bp) the MLV *gag* sequence from CF1 cell RNA was amplified and sequenced and phylogenetic identification was carried out. A positive control for the PCR was 22rv1 cell culture eluate containing XMRV *gag* RNA, using a serial dilution titration, in the two steps of the nested assay (Figure 4.2.1A).

Amplification products in the MLV *gag* gene target area (highly conserved for ecotropic, polytropic and xenotropic sequences alike) are shown in Figure 4.2.1B. CF-1 derived cell-free supernatants were selected from various time points in the 5-IDU and 5-AzaC treatment arms from cell culture (Chapter 3) and from untreated growing cells, particularly samples that had demonstrated the highest RT signal by PERT assay. While one well failed to produce a visible band on Lane 12 (CF1 untreated sample, day 10), the remaining have all amplified an MLV *gag* product of 110bp, at the expected product size. Selected amplicons were purified and sequenced with the second step *gag* reverse and forward primers. The results were confirmed by sequencing of the positive control 22rv1 (Figure 4.2.1B Lane 8). This positive control produced a 110 bp *gag* sequence that mapped in Blast*n*

with 99% pair-wise similarity and E-value 0.0 at complete confidence, with the 22rv1 cell- associated XMRV sequence (GenBank Accession No FN 692043.2).

From sequencing analysis, the CF-1 derived samples exhibited a 110 bp *gag* consensus sequence (Geneious, CLUSTALW) that mapped in a Blastn search with E values close to e^{-116} and pairwise identity of 99.4% with Acc No KJ668270 (ecotropic MLV, full genome), Acc No KJ668269 (Mus musculus DNA strain NOD/ShiLtJ endogenous MLV30) and Acc No KJ668271 (polytropic murine leukaemia virus, complete genome). With an E value of e^{-114} and pairwise identity of 98.9% the same CF1 cell-free supernatant *gag* sequence aligned with Acc No J01998 (AKV provirus) and with an E value of e^{-112} and pairwise identity of 98.5% with Acc No U43202 (RFB provirus) as well as other MLV ecotropic and polytropic sequences (equal or lesser scores).

4.2.2 Further identification of the CF1-related MLV-RNA via MLV *env* PCR

An envelope-specific PCR was developed for its ability to discriminate between ecotropic and xenotropic/polytropic isolates. This is described in detail in the Methods (Section 2.5.4) and sequences are found in Appendix 1. An end-point serial-dilution of 22rv1 cell-free supernatant containing XMRV was carried out. Amplification was possible at the lower detection limit of at least 2500 copies/µl of XMRV RNA (without ultracentrifugation and requiring separate RT synthesis) (Figure 4.2.1C). The end point limit of this conventional PCR approach (2500 RNA copies/µl) was sufficient as it is comparable to the same limit for the *gag* sequencing assay (approx. 2500 RNA copies/µl) for the same reaction conditions. The XMRV DNA plasmid (Vp62) was used as a DNA positive control.

Results are shown in Figure 4.2.1D. A total of 9 CF-1 related samples from peak induction points, the amplification of 5 of which are depicted in the electrophoresis gel in Figure 4.2.1D (Lanes 1-10) were sequenced. The CF-1 untreated sample from day 10 for which MLV *gag* nested was not amplified (Lane

12 in Figure 4.2.1B), produced a 600bp *env* sequence (Lanes 9/10, Figure 4.2.1D) that was included in the analysis.

A consensus sequence was produced (Geneious, CLUSTALW) from all 9 samples of at least 500bp of sequencing product, confirming the presence of one identical MLV-RNA envelope sequence in the CF1 cell-free supernatants. The positive control 22rv1 *env* amplification product was aligned with 99% similarity to the XMRV 22rv1/CWR-R1 complete provirus (Acc No FN 692943.2).

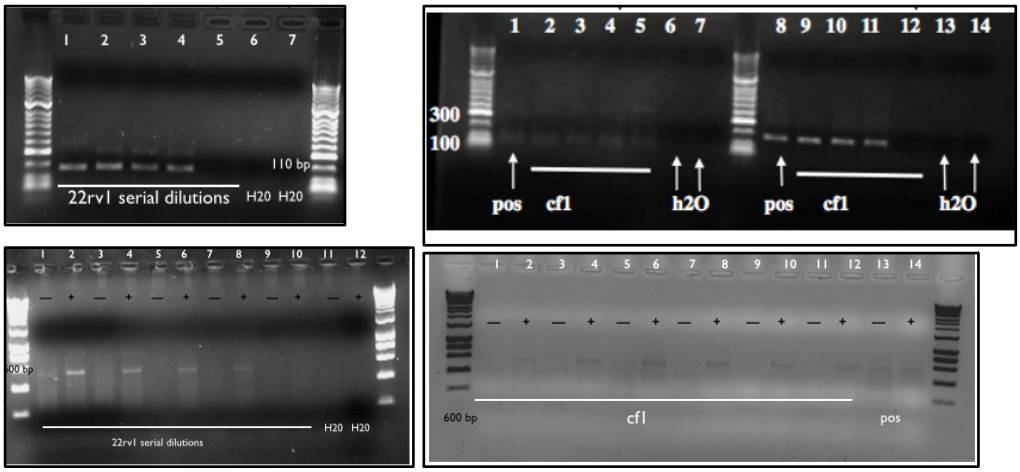
From 18 reverse and forward unassembled sequences, 15/18 directly identified as RFB-MLV (Figure 4.2.1Ea) when analysed in Blast*n, i.e.* aligned with the RFB provirus of MLV (Accession No U43202.1) at the maximum alignment score of 99%. The maximum pairwise identity ratios from 7 fully paired and 4 partial sequences with RFB-MLV ranged from 81%-99% (pairwise identity ratio average: 94.88%, StDev +/- 5.71%) while similarity next identified MLV-AKV (Acc No AF 169256.1) ranging from 81% -98% (pairwise identity ratio average: 93.57% StDev +/- 5.45%) and MLV-SL3-3 with the same E value and pairwise identity as MLV-AKV (not shown). The MLV isolate SL3-3 has 98% similarity with AKV with the exact same envelope sequence. It is a recombinant of AKV with MCF (mink cell foaming) viruses, due to a variation of 200bp within the integrase gene that matches MCF virus sequences (Lund and Pedersen, 1999). For the purposes of clarity, SL3-3 will be considered identical to AKV for analysis purposes, but will be considered for analysis in CF1 cells at a later point during whole genome sequencing.

The derived consensus CF1 related MLV e*nv* and individual envelope sequences were compared with viral sequences previously arising from *gag* gene sequencing. These included RFB, AKV and endogenous ecotropic MLV-30 (emv-30) and polytropic MLV from Nod/ShiLtJ mice (Figure 4.2.1Eb). The polytropic virus failed to align with the sampled products at the envelope region, and was rejected from further analysis (Figure 4.2.1E: top sequence). Amongst AKV, emv-30, RFB and ecotropic MLV the similarity was marked apart from two nucleotide

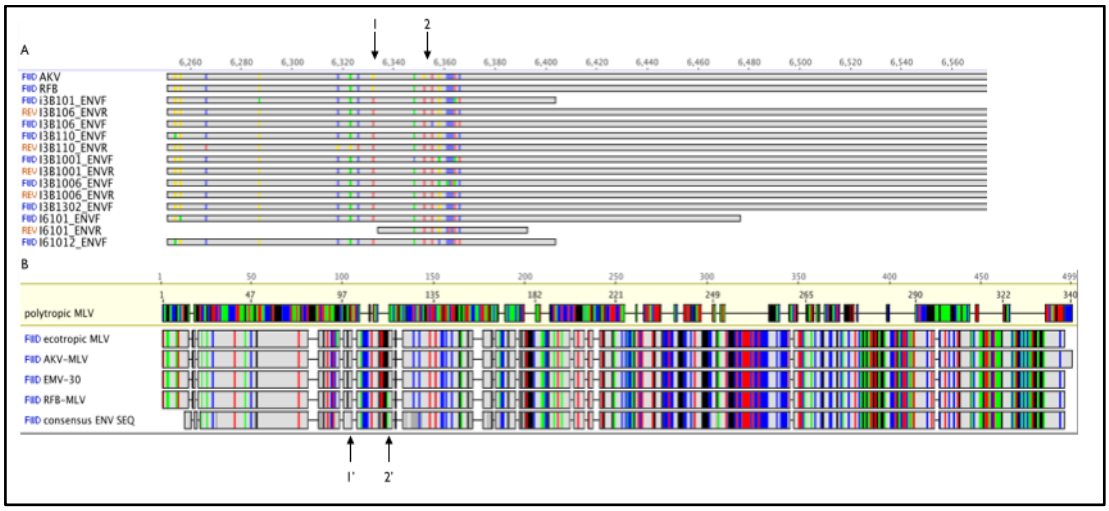
differences (Figure 4.2.1E arrows 1’ and 2’) that are both A/G transitions. A->G substitutions are considered to be the result of the action of APOBEC3 in mouse cells. While at position 1’ this seems to be specific for the CF1 requested sequence, position 2’ indicated RFB-MLV. The alignment could not discriminate further between AKV-MLV and RFB-MLV or emv-30. Both the 1’ and 2’ mutations are synonymous in the translation of the envelope proteins and although they might have secondary structural effects, they are not considered important (Salim et al, 2008).

Figure 4.2.1 Amplification of MLV *gag* and *env* from CF1 cell supernatant

A. B.



C. D.



E.

Figure 4.2.1 Amplification of MLV *gag* and *env* from CF1 cell supernatant

A. **XMRV *gag* nested PCR** 22rv1 cell supernatant eluate (XMRV-RNA) was RT synthesised and 1µl of cDNA was serially diluted (2.5 x10^6 copies/µl -250 copies/µl). The 2nd step amplification products are shown in the gel, at 110bp (Hyperladder II-Bioline, UK). The limiting dilution is found at 250 copies/RNA per µl (Lane 5). Lanes 7-8 are water negative controls.

B. **MLV *gag* nested PCR for CF1 cell supernatant** First-step (lanes 1-7) and second-step (lanes 8-14) show MLV *gag* amplification products for CF1-cell supernatant eluates, the positive control (22rv1 cell) and water negative controls as indicated by arrows. Lane 1/8: 22rv1 supernatant; Lane 2/9: CF1- 5-IDU (50 µg/ml) treated sample, day 1 Lane 3/10: CF1 5-IDU (50 µg/ml) treated sample, day 10, Lane 4/11: CF1 untreated sample, day 1, Lane 5/12; CF1 untreated sample, day 10. Lane 5/12 (CF1 cell sample-untreated) did not amplify, believed due to copy number limitation. Lanes 1/8 (positive control) and 2/9 (CF1 5-IDU treated, day 1) were sequenced directly.

C. **XMRV *env* sequencing PCR** 22rv1 cell-free supernatant (XMRV-RNA) was RT synthesised and 1µl of cDNA was processed in five serial dilutions (corresponding to 2.5 x10^6 copies/µl - 250 copies/µl RNA) in *env* amplification. On the gel is shown the one-step PCR result for the RT-synthesised and non-RT synthesised dilution series and for two negative water controls. As with MLV nested *gag* PCR (Figure 4.2.1A) the same limiting dilution (250 copies/µl) is not efficiently amplified (Lanes 9-10).

D. **MLV *env* sequencing PCR for CF1-cell supernatant** eluates were processed for *env* PCR with and without RT synthesis (lanes 1-10), one negative water control (Lane 11) and a 22rv1 positive control (RT synthesised) (Lane 12). All RT-synthesised samples are amplified. Lane 1/2 CF1 AZA treated (0.5 µg/ml) day 1, Lane 3/4 CF1 AZA treated (2 µg/ml) day 10, Lane 5/6 CF1 5-IDU treated (10 µg/ml) day 1, Lane 7/8 CF1 untreated day 1, Lane 9/10 CF1 untreated day 10. All samples were subsequently sequenced directly post purification.

E. **Alignment of MLV env sequences derived from CF1 cell supernatant** Shown on top of graphic A: I3101 to I3B1301 are 13 unassembled sequences corresponding to cell culture induction days (1-13) from the PCR reaction in 4.2.1D and aligned individually with the two main viral sequences AKV and RFB-MLV. Bottom of the graphic (B) depicts the CLUSTALW alignment of the sample consensus sequences derived from A. The polytropic envelope sequence (first in bottom alignment) does not align. Xenotropic sequences have a significant gap in alignment and were discarded from the alignment. Arrows in positions 1 (1') and 2 (2') demonstrate the two main (>95%) nucleotide differences in the CF1 related *env* sequence consensus and AKV, RFB or emv-30.

4.2.3 Phylogenetic analysis of *env* gene derived sequences

A phylogenetic tree was constructed from envelope (*env*) gene sequences. The analysis further allowed for comparison to the two osteosarcoma viruses present in CF1 mice (FBJ and FBR), with other xenotropic and polytropic MLVs as well as ecotropic MLVs and AKV. The tree is shown in Figure 4.2.3.

Figure 4.2.3 Phylogenetic tree of *env* sequence consensus from CF1 cell supernatant

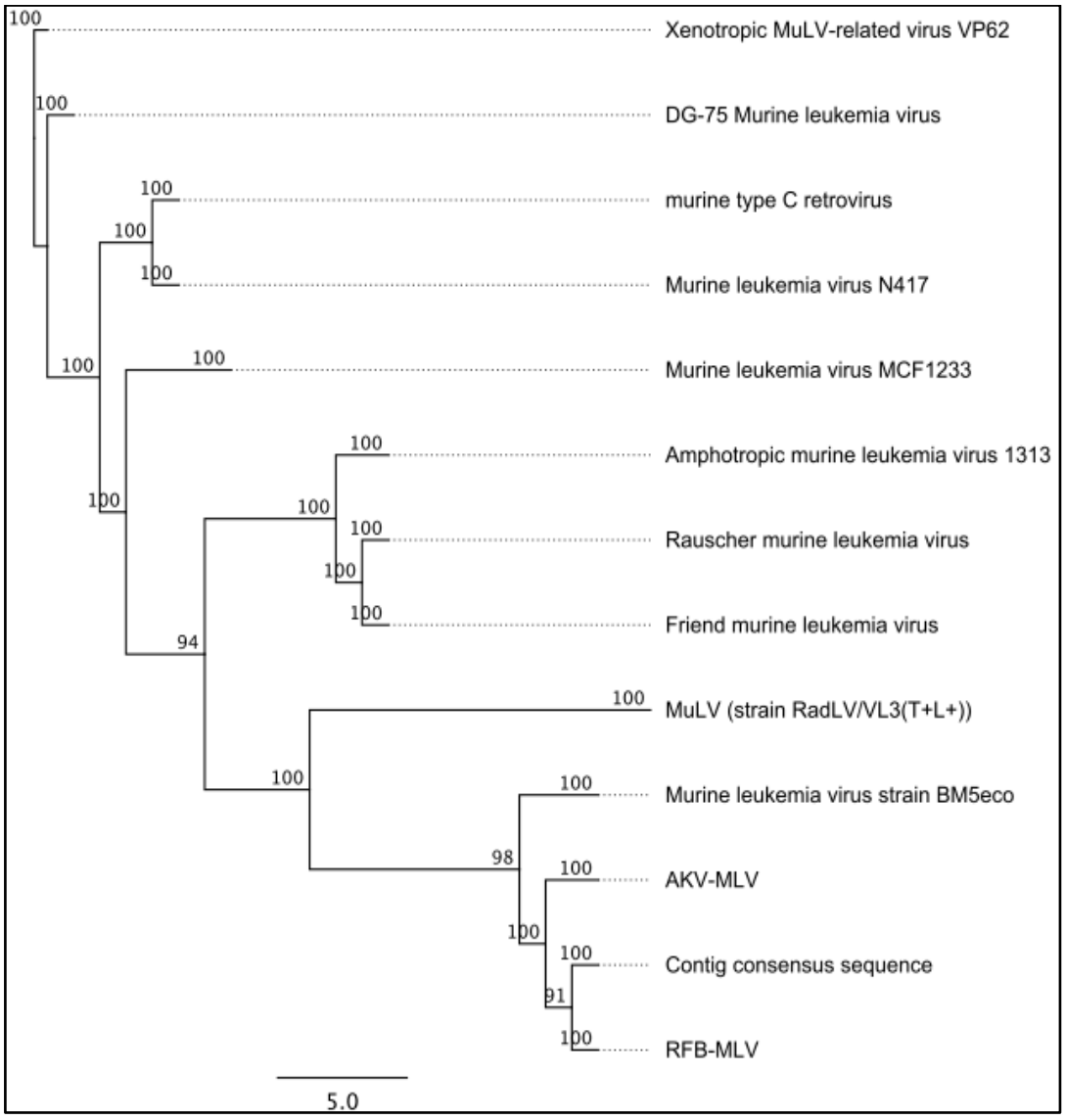


Figure 4.2.3: Phylogenetic tree of *env* sequence consensus from CF1 cell supernatant

A phylogenetic tree was constructed to measure base substitution distance between *env* sequencing consensus from CF1 cells (contig consensus) and a large number of MLV isolates. The tree is constructed using Geneious (version R87, Bio) using the Neighbor-Joining method (Jukes-Cantor) and Bootstrap re-sampling. Branches represent distance according to base substitution ratios which is a measure of paired sequence differences to the single nucleotide level (scale bar); while they are also labelled for consensus support (%) (numbers on branches).

4.2.4 Immunofluorescence staining for MLV Gag in CF1 cells

Immunofluorescence (IF) staining for anti-MLV Gag (Alexa-488) was performed according to a standard protocol described in detail in Methods 2.6.1. The cells were tested for expression of the Gag protein as shown in Figure 4.2.4. Cells that are expressing MLV Gag have a characteristic granulate punctuate green fluorescence at positions that are mainly cytoplasmic but also peri-nuclear. The 22rv1 cells, that have a high titre of expressed XMRV at 10^8 copies/µl are significantly denser in fluorescence. The CF1 cells are also distinctively positive while other mouse cells such as the clinically- exploited J2 MEF cell line are negative. A human negative control (293T cells) gave no IF signal.

Figure 4.2.4 Immunofluorescent staining of cells containing MLV Gag

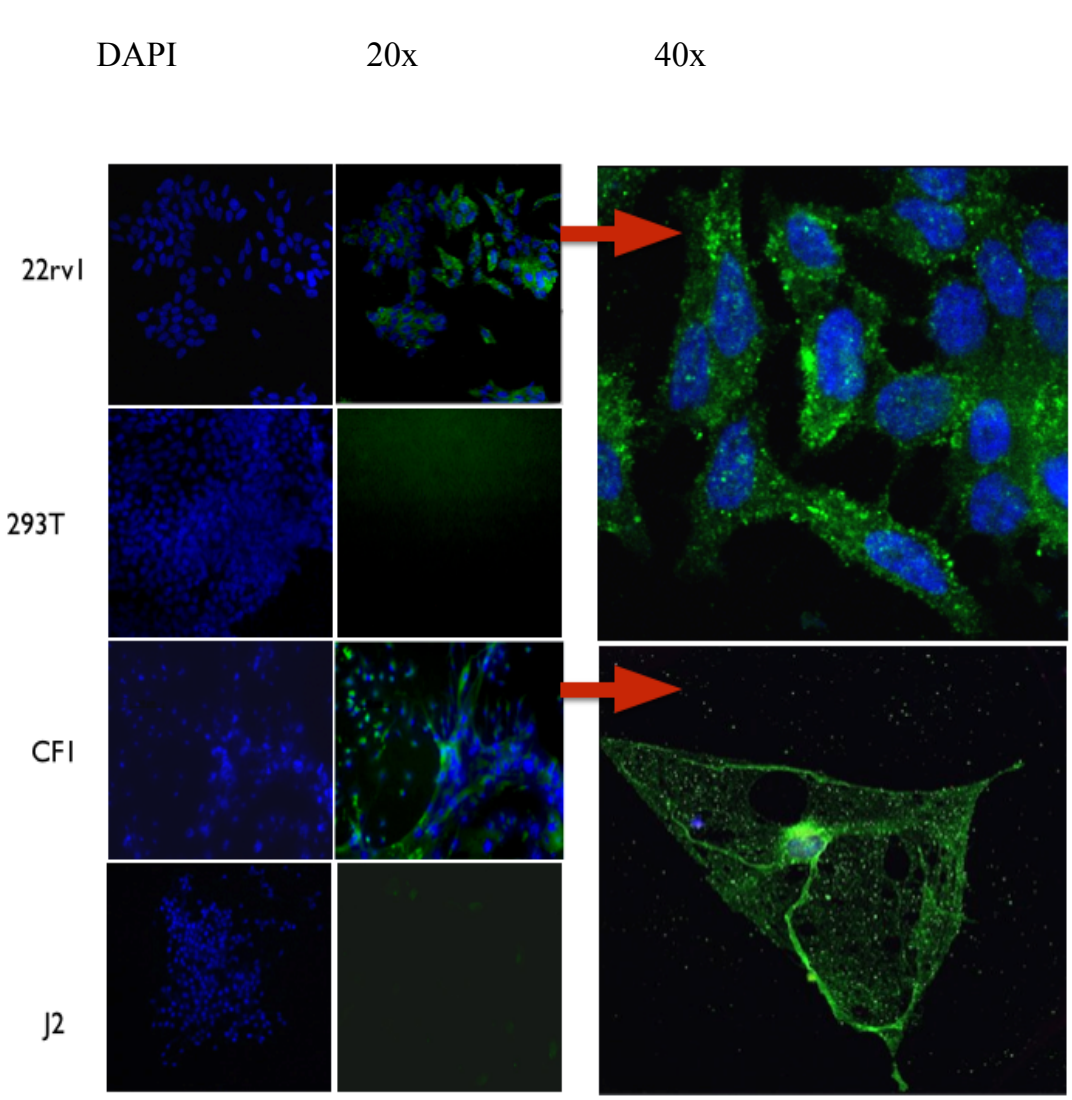


Figure 4.2.4. Immunofluorescent staining of cells containing MLV Gag

Top row is 22rv1 cells (from left to right): DAPI nucleolar staining (blue), Anti-MLV GAG (green) at magnifications 20x and 40x. Second row is the negative control - a human embryonic kidney 293T cell line. Third row is CF1 cells (from left to right) in a DAPI nucleolar stain, magnification 20x and 40x. Bottom row is J2 mouse embryonic fibroblast cells that are negative. Cells grown on 9mm glass coverslips were fixed (4% PFA) and permeabilised with saponin 0.02%. Cells were then stained with 1:5000 of anti-*GAG* (Abcam) and 5µg/ml of secondary, or primary antibody alone.

4.2.5 Distinction of RFB-MLV and AKV-MLV via tandem repeat LTR PCR

CF1-cell related MLV is similar to AKV MLV and retroviruses belonging to its class (such as SL3-3 or murine sarcoma viruses). A PCR was designed using an AKV-specific primer pair to answer the question whether AKV-MLV, which is highly leukaemogenic was present in CF1 cells (See Methods 2.5.4).

Identifying the U3-LTR as the only target available for the discrimination of the two isolates, the primers were directed at the one unique region that is AKV-specific. The AKV U3 bears duplications or tandem repeats comprising of short identical 20-50bp sequences that are enhancers of oncogenesis, giving the virus its high leukaemogenic ability. RFB also has duplications in the same region, however it bears differences in their organisation. These are graphed in detail in Figure 4.2.5. Additionally, in RFB-MLV there are three base pair substitutions in positions 65,157 (repeated in 221, 285) and 181 (repeated in position 245) and a deletion in position 345 and 346 (where an extra C is found in AKV) (Strauss, 1992) for the LTR region (not shown). Two tandem repeats unique to AKV were identified as best primer positions namely BC and EB. Forward and reverse primers in this gene area were designed and multiplexed with available primers for generic MLV- LTR regions, namely U3_LTR reverse.

The sequences amplified in the tandem repeat PCR are seen in figure 4.2.5 aligned against AKV and RFB-MLV. As a positive control for AKV-MLV integrated provirus, genomic DNA from AKR-2B cells was used, known to contain multiple copies of AKV-MLV. An ecotropic virus-negative mouse genomic DNA control (B6 mouse gDNA) was used as a negative control. A water negative control was included (not shown). From post-PCR product purification and direct-sequencing analysis, AKR-2B gDNA resulted in a 265bp sequence with 99% identity to the AKV-MLV LTR region (Accession No J01998). For CF1 gDNA, the resulting 265bp sequence also aligned to 99% with AKV-MLV LTR. Subsequent analysis confirmed that a clear AKV-like tandem repeat pattern (i.e. BCDE-BCDE-U3), and not RFB-MLV tandem (ie BE-CDE-CDE-U3) was present in CF1 genomic DNA.

Figure 4.2.5 shows the analysis of the derived sequences that are identical, either from AKR-2B genomic DNA (products 1- 2 and 3) or for CF1 gDNA (products 4 and 6). The possibility of a product arising from a mis-annealing of the BC primer, on a pattern of tandem repeats similar to that observed in RFB-MLV i.e. B-ECD-ECD-E is faint; as the E tandem is only 15bp long it would take a secondary hairpin structure to cause a miss in the sequential amplification of base pairs- but that would be expected only in the first tandem repeat and not the second that lies closest to the U3; the lack of an intermediate E tandem repeat on the products obtained from CF1-MLV is indeed confirmed by the presence of U3 on the right direction of the BC primer binding site in the two obtained sequences. Both forward sequences from CF1 gDNA are completely concordant with the pattern observed in AKV i.e. BCDE-U3.

Interestingly, comparing with the unique sequences available in the Nucleotide database, for two osteosarcoma-inducing CF1 mouse-derived viruses, one derived from an integrated 3.8Kb FBR-MSV provirus in *ratus norvegicus* (Acc No K02712) and a second 4.2Kb FBJ-MSV proviral sequence (Acc No V01184.1) from CF1 mice, it is evident that both these osteosarcoma viruses from CF1 mice bear one full repeat from the AKV tandem pattern i.e. BCDE. Amongst the two MSVs, the CF1-derived 265bp sequence aligns in symphony with FBR rather than FBJ-MSV, as the latter possesses a second repeat area within the amplified region.

This is a new analysis of the tandem repeats of the CF1-MLV virus family LTR regions. These results agree with the previously described presence of the AKV-like sarcoma FBR virus in the genome of CF1 cells. They do not however further help to distinguish between AKV and RFB. This question is now addressed via full length sequencing of RNA viral content and MLV integration analysis by next generation sequencing.

Figure 4.2.5 Tandem repeat MLV-LTR PCR for AKR gDNA and CF1 gDNA



Figure 4.2.5 Tandem repeat MLV-LTR PCR for AKR gDNA and CF1 gDNA

The alignment shows the tandem repeats organisation in the viruses shown (original work) in line of appearance RFB-MLV, FBR-MSV, FBJ-MSV and AKV-MLV. Tandem repeat PCR products are from AKR genomic DNA (products 1,2,3) and for CF1 genomic DNA (products 4 and 6).The derived sequences are shown for products 1,2,3,4 and 6 in duplicate (forward and reverse orientations). There is complete alignment of the orientations with the AKV-like tandem repeats but not for RFB, FBR or FBJ viruses.

4.2.6 Syncytia formation assays on NIH-3T3 cells

RFB-MLV is reported to be infectious for mice and mouse cells, causing osteoma and osteopetroma in inoculated mice (Pedersen et al, 1992), hence the CF1 MEF cell derived virus was assessed for its cytopathic effect on NIH-3T3 mouse and on XC rat cells for the appearance of characteristic multi-nucleated syncytia (see Methods Section 2.4.1).

Infectivity tested was investigated by a. inoculation of potentially infected cell supernatant (CF1 cells and J2 cells) and by b. co-cultivation of infected cells (CF1 cells and J2 cells). A positive control for the appearance and enumeration of syncytia was the Balb-K cell line (supernatant and cells) known to express helper N-tropic MLV and Kirsten MSV (Aaronson, 1971). The presence of enlarged multinucleate syncytia that form due to viral infection was monitored over 7 days via light microscopy and were assessed as previously described in McClure et al, 1990. The assay used a serial dilution method (10-fold dilutions) to assess cytopathic effect at each dilution over supernatant. Observed syncytia were stained with DAPI (nuclear) and lectin (cytoplasmic).

Supernatant used was from untreated and treated cells and the assay was performed on NIH-3T3 cells grown to confluence for 48h. Serial dilutions ranging from 10^{-1} to 10^{-4} of the inoculum were tested in duplicate. Two 3T3 wells served as negative controls. Supernatant from the tested cells was passed serially onto another layer of uninfected 3T3 cells for a subsequent 7d. Similar incubation and growth conditions were applied for syncytial formation tested via co-cultivation on NIH-3T3 cells, albeit the serial dilution method was not applicable.

Treated and untreated CF1 cell supernatant infected equally 3T3 mouse cells with the appearance of syncytia at 3 days post-inoculation (Figure 4.2.6E) while the number of syncytia remained constant for 7 days. J2 cell derived supernatants did not form syncytia (not shown).

Supernatant from syncytial assays serially passed onto uninfected NIH-3T3 cells did not produce new syncytia as monitored for an extra 7 days. Indeed, for Balb-K supernatant, in similar infection conditions, MSV defective particles have been shown to lack any onward infectivity (Bassin et al, 1971). It is possible that the MLVs in the CF1 and Balb-K supernatant stocks are equally unable to productively infect NIH-3T3 cells, through post-entry FV-1 restriction. Supernatant from CF1 cells also does not produce transferrable cytopathic effect when tested via supernatant inoculation on our virus-negative J2 3T3 cells, that are also N-tropic (data not shown).

For the co-cultivation of CF1 and NIH-3T3 cells, CF1 cells initially produced syncytia the number of which reduced considerably within 7 days (Table 4.2.6:2) contrary to Balb-K cell and NIH-3T3 cell co-cultivations that showed a constant number of syncytia for 7 days. When supernatant from the co-cultivation experiments was further passed onto a new layer of uninfected NIH- 3T3 cells no syncytia were produced both for CF1 - 3T3 and Balb-K- 3T3 co-cultivations.

These findings are consistent with the published observation that CF1 mouse related MLV i.e. the historic RFB-MLV does not produce a productive cytopathic effect in mouse cells (Pedersen et al, 1992). In the case of Balb-K cells, the MLV identified is also not able to produce a cytopathic effect, as it is unable to efficiently propagate as a replication competent virus.

TABLE 4.2.6 Syncytia formation assays via co-cultivation

passage 1				passage 2	
dpcc*	3	5	7	3	7
4 replicate average/ cell type					
CF1	+++	++	+	-	-
BALB-K	+++	+++	+++	-	-
NC	-	-	-	-	-
Key	(+++) >25% of nuclei in syncytia	(++) 5-25% nuclei in syncytia	(+) 1-5% nuclei in syncytia	(-) no discernible syncytia formation	(McClure et al, 1990)
*dpcc=days post cocultivation					

TABLE 4.2.6 Syncytia formation assays via co-cultivation NIH-3T3 cells (ECAAC) (3x10^4 cells/well) were seeded in co-cultivation with untreated Balb-K cells (10^5 cells/well) and with untreated CF1 (10^5 cells/well) in 6-well plates in duplicate. The co-cultivated cells were observed for 7 days for the presentation of cytopathic effect (syncytia formation), indicative of a productive retroviral infection. The table shows the enumeration of syncytia per co-cultivation well across 7 days for passages 1 and 2. Quantification is based on the average of 10 visual fields per well, observed in x20 magnification via direct microscopy. The scale of assessment was based on McClure et al, 1990.

Figure 4.2.6 Syncytia formation assay for infectivity assessment on NIH-3T3 cells

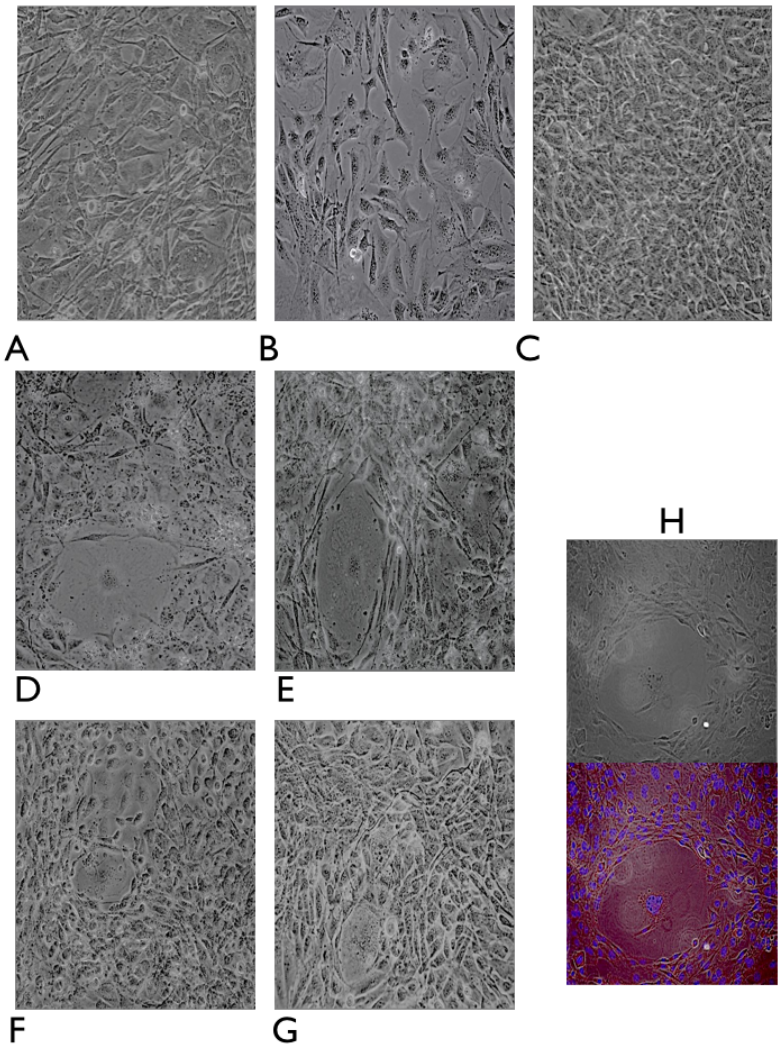


Figure 4.2.6: Syncytia formation assay for infectivity assessment on NIH-3T3 cells

Light microscopy of cell lines tested via supernatant or co-cultivation for cytopathic effect (syncytia formation) on NIH-3T3 cells (20x magnification). In detail, A) Balb-K positive control cells (untreated) B) CF1 cells (untreated) C) NIH-3T3 cells D) Example syncytium via supernatant inoculation from treated Balb-K cells on confluent NIH-3T3s E) Example syncytium via supernatant inoculation from treated CF1 cells on confluent NIH-3T3s F) Syncytia via co-cultivation of Balb-K and NIH-3T3 cells G) Syncytia via co-cultivation of CF1 and NIH-3T3 cells. H) Immunofluorescent staining of a syncytium formed on NIH-3T3 cells by Balb-K (treated with 5-AzaC) supernatant. Blue: DAPI nuclear stain and Red: lectin cytoplasmic stain (20x magnification).

4.2.7 Direct infectivity assays by syncytia formation on XC cells

CF1-related RFB-MLV virus does not produce a cytopathic effect in NIH-3T3 cells but reportedly can form syncytial plaques on XC- cells (Pedersen et al, 1992). The CF1 cells produced visible giant syncytia on day 1 post-inoculation; with a similar effect when CF1 supernatant infection was assayed (See Methods, 2.4.2 for the modifications of the assay). The syncytia were present at 7 dpi, while uninfected XC controls remained consistently syncytium-free throughout the experiment. A negative virus fusion control, using 22rv1 cell derived XMRV-containing supernatant only produced a sparse cytopathic effect.

Figure 4.2.7 XC cell syncytial formation assay for CF1 cells

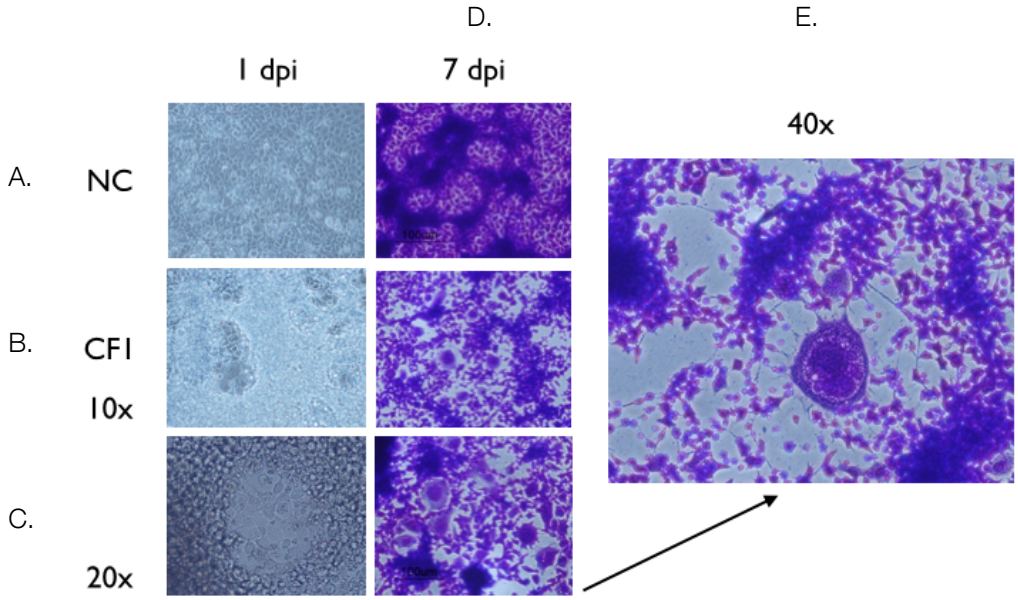


Figure 4.2.7. XC cell syncytial formation assay for CF1 cells

Ecotropic MLV syncytia assay was performed using the XC cell line *rattus norvegicus*. The assay was modified for filtered cell-free supernatant directly inoculated on XC cells. In the figure A. the negative control (NC) uninfected rat cells does not exhibit syncytia whereas in B. CF1 cell supernatant assayed on XC cells demonstrates syncytia at day 1 and day 7 post inoculation. Magnifications of 10x (A,B), 20x (C) and 40x (E) are seen. The plaques are fixed with methanol and additionally stained with 0.05% crystal violet at 7 dpi (D).

4.2.8 Spliced RFB-MLV envelope mRNA is transcribed in CF1 cells

A spliced *envelope* real time PCR was designed in-house to detect active transcription of envelope mRNA of CF1-MLV virus in growing CF1 cells. The design of the primers for the envelope spliced RNA and the RT-PCR conditions are based on a conserved region of *env* and described in Methods 2.5.4 and Methods 2.5.8.

The input RNA was tested in serial ten-fold dilutions, in three replicates per dilution and the negative control consisted of total RNA extracted from growing J2 mouse fibroblast cells. A transcription control, the mouse HPRT gene was amplified consistently from CF1 and J2 cell RNA serial dilutions, allowing for the construction of a standard curve with a slope of -3.37 and R2 of 0.991 (Figure 4.2.8A). Replicates showed minimal variation per dilution triplicate (STDEV ±0.25 cycles). Quantification output was in mRNA copies/reaction.

Messenger RNA RT-PCR was positive for dilutions of total RNA from CF1 cells. Average CF1-*m*RNA *env* copies were calculated in triplicate against the quantification mHPRT curve (Figure 4.2.84A) and with the ΔΔCT (Differential Cycle Threshold) method. On average, the mRNA *env* content of untreated growing CF1 cells was amplified 4.02 CT earlier ±1.46 CT (STDEV) for three replicates per serial dilution than HPRT equivalents. The ΔΔCT is calculated as the difference between CT(HPRT)-CT(mRNA *env*), suggesting that the *envelope* mRNA signal is being expressed at least 16-fold higher compared to the stably expressed HPRT mRNA (ΔΔCT expression=$2^{-\Delta\Delta_{CT}}$). The base of 2 assumes 100% qPCR amplification efficiency for all reactions, or a doubling of amplicon with each subsequent qPCR cycle, a condition which is met as equal PCR efficiencies of 100% for both targets are being measured (PCR efficiency for HPRT=100.5%, slope -3.310 and PCR efficiency for mRNA *env*= 99.8%, slope -3.327).

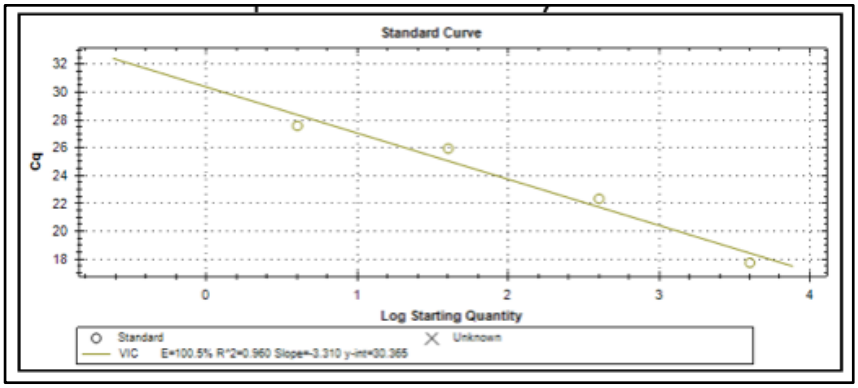
The input RNA was also quantified to control for corresponding cell number. As one cell yields approximately 6pg of total RNA, our standard curve of 25ng-25pg is also equivalent to 4,150 to 4.15 cells /reaction. The results of this extrapolation suggest that

at least 16.2 mRNA *env* copies (±2 copies average STDEV) are expressed per growing CF1 cell. However, this depends on the assumption that HPRT is a single copy reference gene. The maximum expression of HPRT is 10 copies/cell, suggesting that CF1 mRNA *env* maximum expression is 162 copies/cell (±20 mRNA *env* copies STDEV). This expression corresponds to the observed MLV *gag* RNA content of CF1 supernatant as calculated in real time PCR that corresponds to 32-40 copies/cell in treated and untreated cells on average (Chapter 3). As hypothesised, J2 cells did not amplify positive in the mRNA *env* RT-PCR and the NTC remained negative.

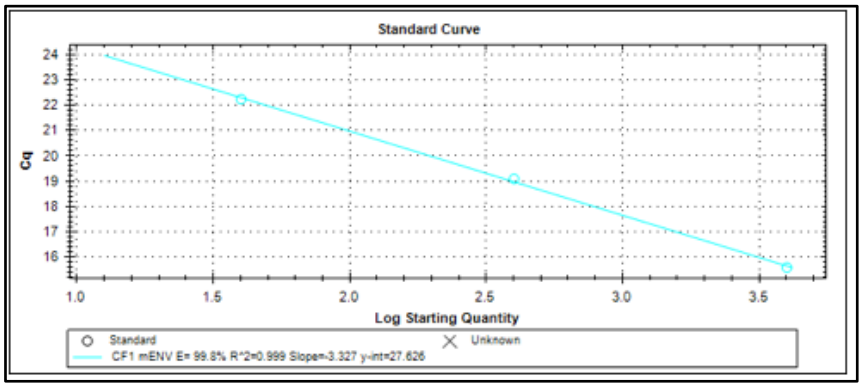
Finally, in order to confirm that the spliced *envelope* PCR amplified the intended target, one random product from the mENV amplification of the CF1 total RNA was sequenced directly. The target 5'-LTR start region (bases 598-648 in the RFB-MLV genome) and envelope gene tail (bases 6117-6417 in the RFB-MLV genome) aligned 100% with our spliced envelope RT-PCR product.

Figure 4.2.8 Transcription of spliced messenger *env* RNA from CF1 cells

A.



B.



C.

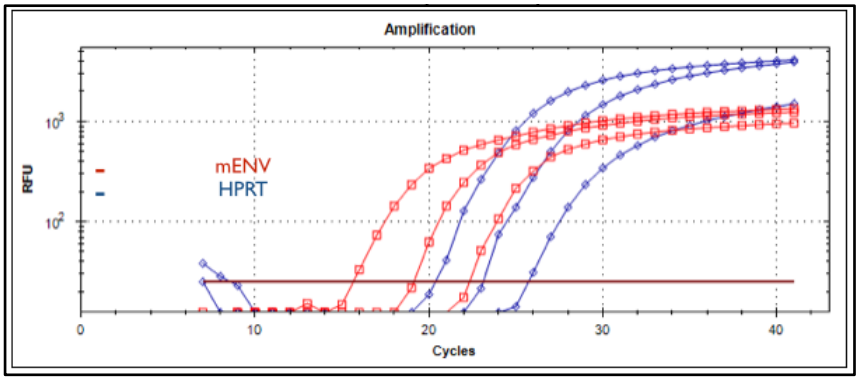


Figure 4.2.8. Transcription of spliced messenger *env* RNA from CF1 cells

A. The standard curve of the stably expressed mouse HPRT mRNA reference control in CF1 total cell RNA: Ten-fold dilutions of total RNA from CF1 cells were amplified for mouse HPRT mRNA. The standard curve has a slope of -3.310 and R2 of 0.96, with an efficiency of 100% for 2-fold exponential amplification.

B. The amplification of messenger envelope RNA in CF1 total cell RNA: Ten-fold dilutions of supernatant from CF1 cells were amplified for spliced envelope mRNA. The dilutions produce a curve that has a slope of -3.320, R2 of 0.99 and an efficiency of 99.8% for 2-fold exponential amplification. This curve is constructed to demonstrate comparative efficiency in the amplification of the target gene versus the reference gene control.

C. Example fluorescent amplification curves of target and reference gene in CF1 total cell RNA: Red lines represent spliced envelope RNA amplification and Blue lines represent mouse HPRT gene amplification. The dilutions depicted represent input RNA values of 25 ng to 0.25 ng, that correspond to 4000-40 CF1 cells approximately.

4.2.9. Control for Mitomycin C treatment effect on MLV transcription

MLV RNA novel transcription was assessed in CF1 feeder cells after treatment with various concentrations of Mitomycin C for varying times. We demonstrate that the most frequently used combinations of Mitomycin C treatments for feeder cell line inactivation does not alter the expression of MLV *gag* (Figure 4.2.9) in CF1 MEF cells and cell supernatant over two 6-hour windows within 24 hours.

Figure 4.2.9 MLV *gag* RNA production in Mitomycin C treated CF1 cells

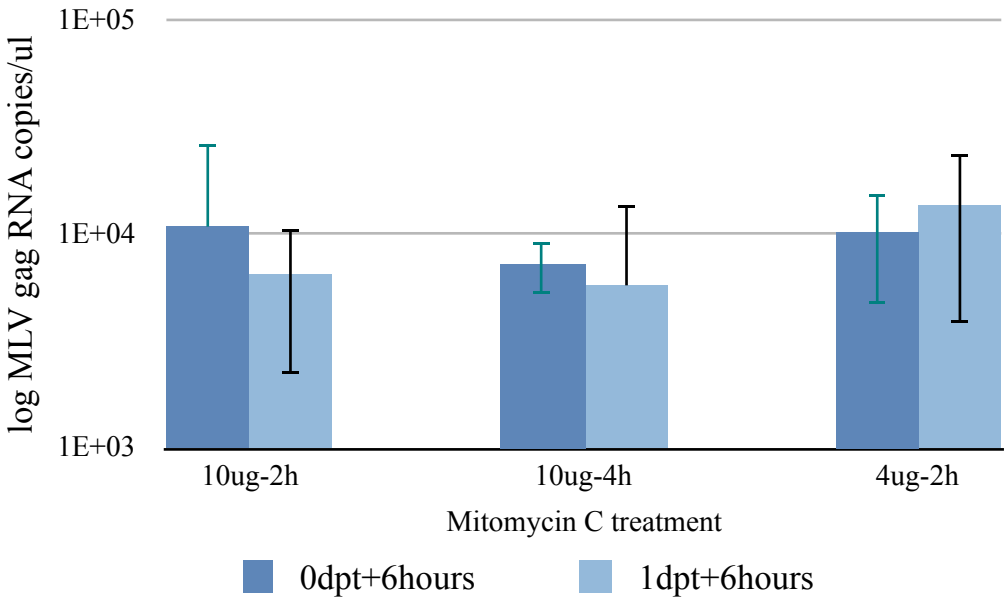


Figure 4.2.9 MLV *gag* RNA production in Mitomycin C treated CF1 cells
Two concentrations of Mitomycin C (10µg and 4µg/ml) were plated for 2 and 4 hours and MLV *gag* RNA was quantified over two six hour periods at post treatment days 0 and 1. RT-PCR quantification was performed for MLV *gag* RNA using ΔΔCT. The ability of the cells to produce MLV *gag* was found to be unaffected by Mitomycin C treatment. Untreated cells have low MLV *gag* RNA content <1000 copies/µl (not shown).

4.2.10 CF1-related MLV virus is unable to infect human DERSE iGFP cells

The virus from CF1 cells was tested on the indicator cell line assay using DERSE iGFP cells. The principle behind the function of the engineered iGFP plasmid is described in detail in Methods 2.6.2. Supernatant from growing 22rv1 cells, containing XMRV was used for the infection of DERSE cells in 3-fold dilutions as a positive control. Triple concentration of supernatant from growing CF1 cells, containing CF1-MLV virus was tested for DERSE cell infection in 3-fold dilutions.

DERSE cells were assessed at 4h, 8h, 1day and each day after that until 15 days post infection (dpi) under a fluorescent microscope for the demonstration of GFP positive cells. DERSE -iGFP LnCap cells are lacking the mCAT-1 receptor for ecotropic retroviral entry. Fluorescent cells producing iGFP were detected at 6 dpi only for XMRV infection, while CF1 infected cells remained negative for 15 dpi (Figure 4.2.10.1). No fluorescent cells were seen in negative uninfected controls.

To control for background fluorescence and monitor the infection of DERSE cells by XMRV and/or CF1-related MLV before the emergence of iGFP fluorescence, infected DERSE cells were stained with anti-MLV Gag one and 2 days post-infection. Indeed, the process of iGFP transcription is expected to take longer than 3 days, since a second cycle of cell infection is necessary for the inverted iGFP genes to be properly transcribed (see Methods Figure 2.6). The results of this immunofluorescence at 48h post infection are shown in Figure 4.2.10.2.

Figure 4.2.10.1 Reporter DERSE cells post-infection with cell supernatant

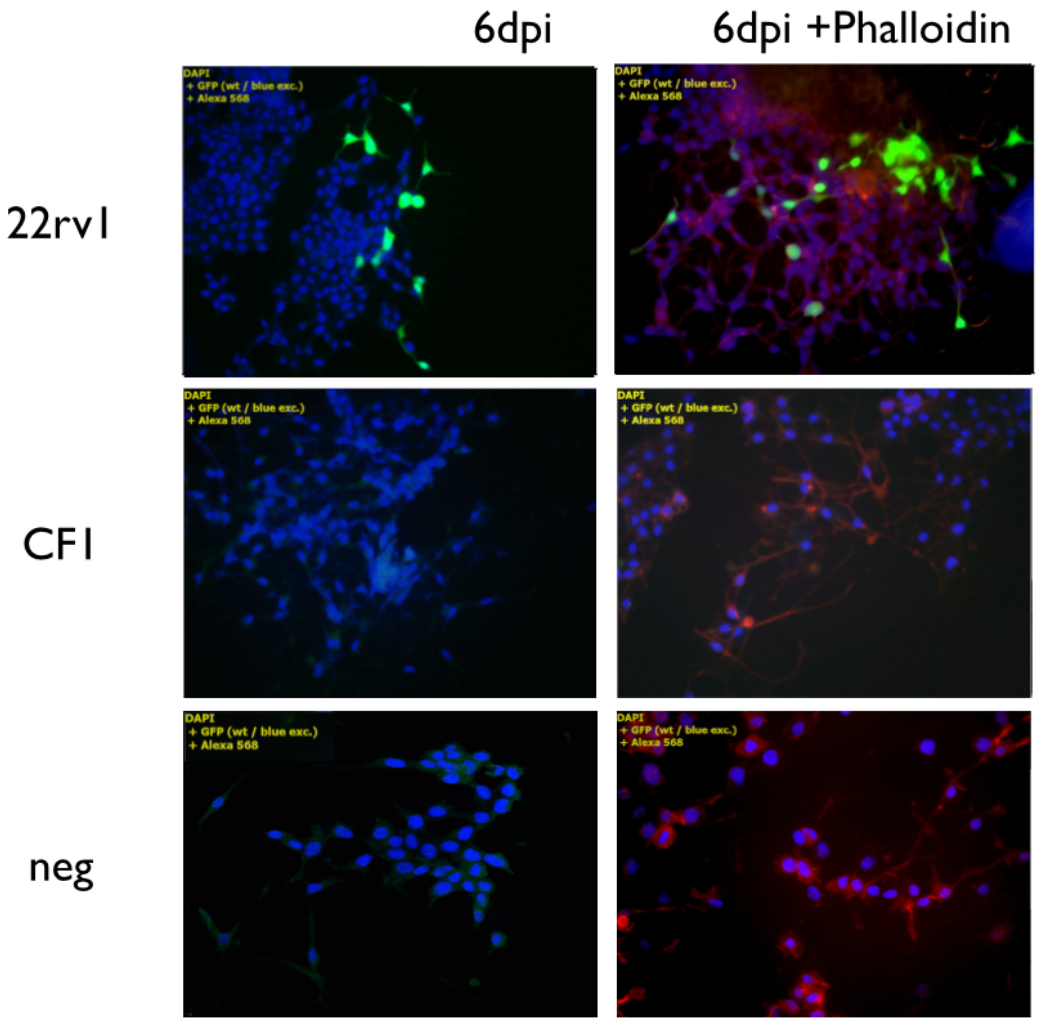


Figure 4.2.10.1 Reporter DERSE cells post-infection with cell supernatant

DERSE iGFP LnCap cells were infected with XMRV and CF1-MLV containing supernatant. **Top** row: XMRV infection (30µl inoculate from 22rv1 cells in 2ml of growth medium/well) produces iGFP positive cells (bright green) at 6 dpi. The fluorescent cells are visualised with and without phalloidin staining, which is specific for eukaryotic cytoskeleton. **Middle row**: CF1 cell supernatant infection (90µl inoculate in 2ml of growth medium/well) that remains negative at 6 dpi and 15 dpi (not shown). **Bottom row**: uninfected negative controls. [Phalloidin is detectable at 568nm (red) and iGFP at 488nm (green)]. Negative controls consisted of uninfected cells and no Phalloidin controls.

Figure 4.2.10.2 The visualisation of MLV infection of DERSE cells prior iGFP emergence with anti-MLV Gag

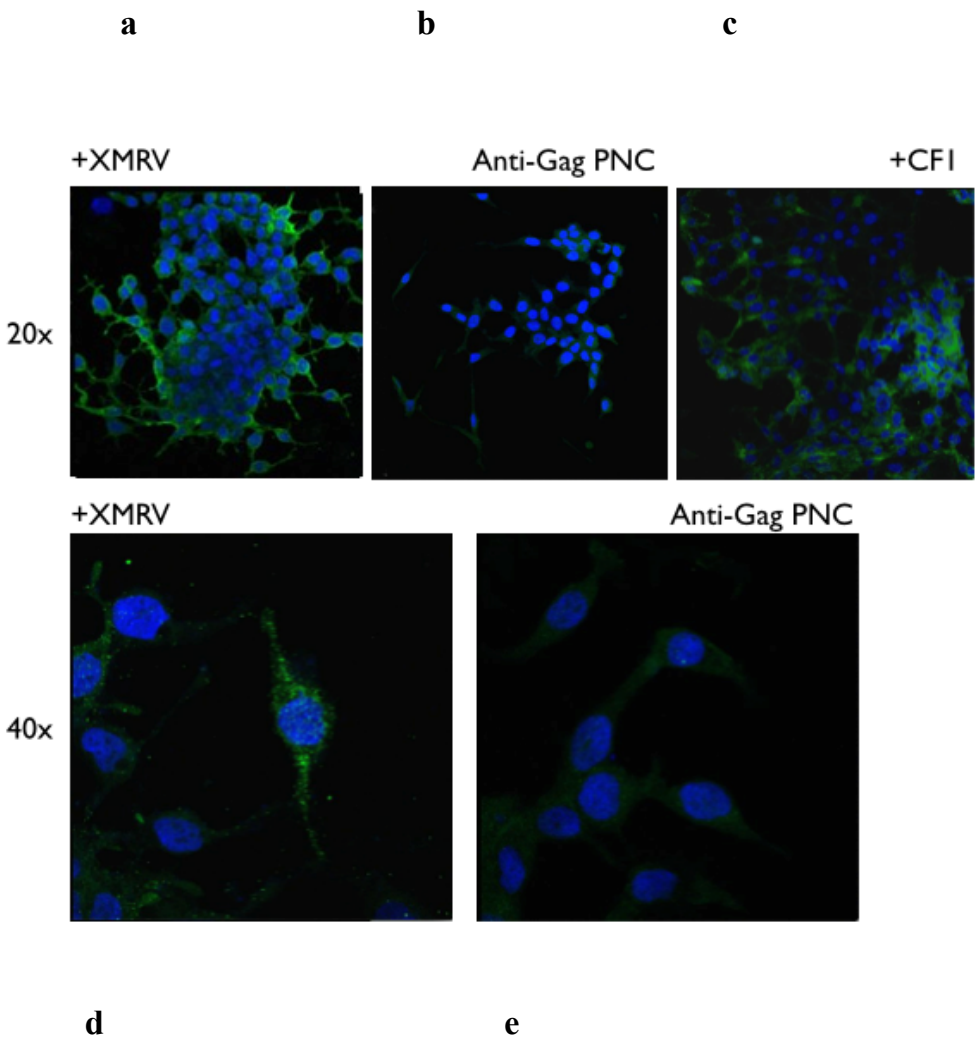


Figure 4.2.10.2 The visualisation of MLV infection of DERSE cells prior iGFP emergence with anti-MLV Gag

Infected (48h post infection) and uninfected cells DERSE LnCap were immunostained with Anti-MLV Gag (Green-excitable at 488nm). XMRV and CF1-MLV infected LnCap cells are both positive for MLV Gag 48h post infection (a and b) while the antibody localises in the *extracellular* environment. Infection with XMRV is evident at the single-cell *intracellular* level with the higher magnification 40x (image d). No pronounced *intracellular* staining is seen in CF1-MLV DERSE cell infection (image c). The experiment used an Anti-Gag primary antibody deficient negative control (Anti-Gag PNC) (images b and e). A low level of autofluorescence is observed at 20x and 40x magnifications (anti-Gag PNC) of infected (shown-images b and e) and uninfected (not shown) controls.

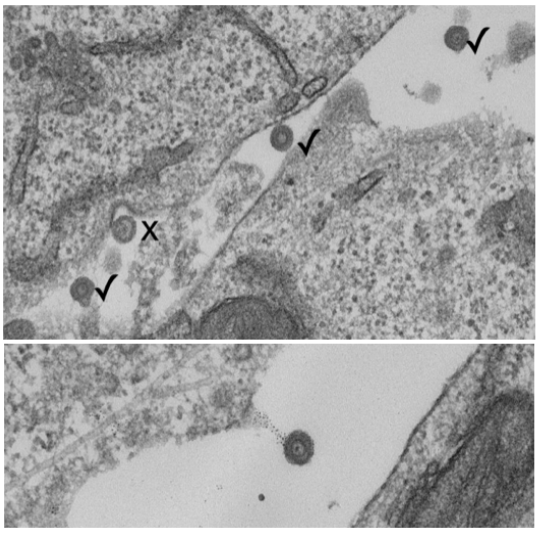
4.2.11 Transmission Electron Microscopy (TEM)

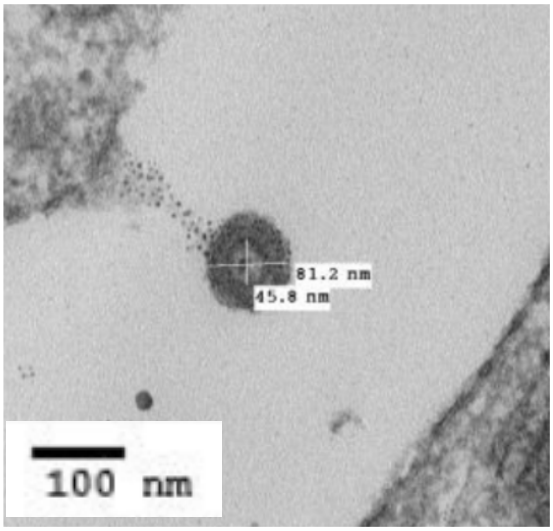
Transmission electron microscopic sections of murine and human cells were obtained following a protocol described in Methods 2.6.3. Two cell lines, the human prostate cancer 22rv1 that is actively expressing XMRV and the CF1-MEF cells, were analysed for virion content at high magnification. Virions that were visualised were scrutinised for morphology, against internal cellular structures such as cell vacuoles. The XMRV virions were identified and found to have a typical retrovirion structure, a total diameter of 80-82nm with a dense capsule (Type C morphology) (Figure 4.2.11 a-b-c). Interestingly most of the particles observed were not dense at the core, suggesting they are immature but several demonstrated significant maturations with condensing cores. Virions were identical in all 22rv1 samples.

CF1-MEF cells were scanned by TEM for virions (Figure 4.4.11 d-e-f). A large number of aggregates of what were closer identified as virions was discovered across sections of CF1 cells. These are mainly located extracellularly, but several sections contained virions at sites within the membrane, whilst a budding virion was also observed. These CF1 virions are not identical in diameter, ranging from 74-97nm while the most important observation is the lack of a capsular hypodensity to suggest a typical mature retroviral core. As depicted in Figure 4.2.11f the virions possess a capsular ring that is closer to the surface suggesting the virions are immature. It is known that mouse cells bear a high copy number of IAP sequences. Although these wild-type elements have type A retroviral morphology they are traditionally confined to the cisternae of the Golgi apparatus and are predominantly located intracellularly.

Figure 4.2.11 Transmission Electron Microscopy of human and mouse cell lines

a.

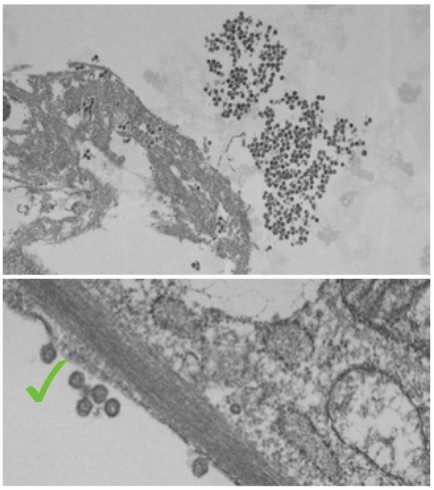


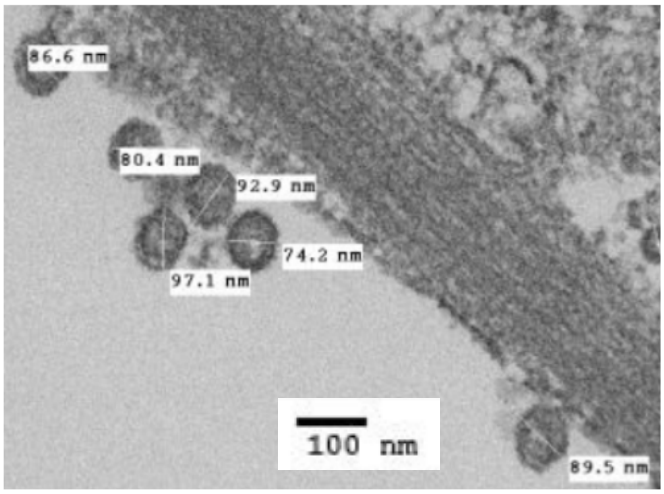


b.

c.

d.





e.

f.

Figure 4.2.11: Transmission Electron Microscopy of human and mouse cell lines

a-b-c. Sections of 22rv1 cells processed for transmission electron microscopy in 25000x direct magnification. The cells were processed and stained using 1% aqueous osmium tetroxide (OsO4) and uranyl acetate. Tick marks indicate morphologies that are virion type, whereas an x mark indicates the structure not to be considered virions (a). A budding XMRV virion is seen in (b). The diameter of the virions is consistent and measures 81.2 nm (c).

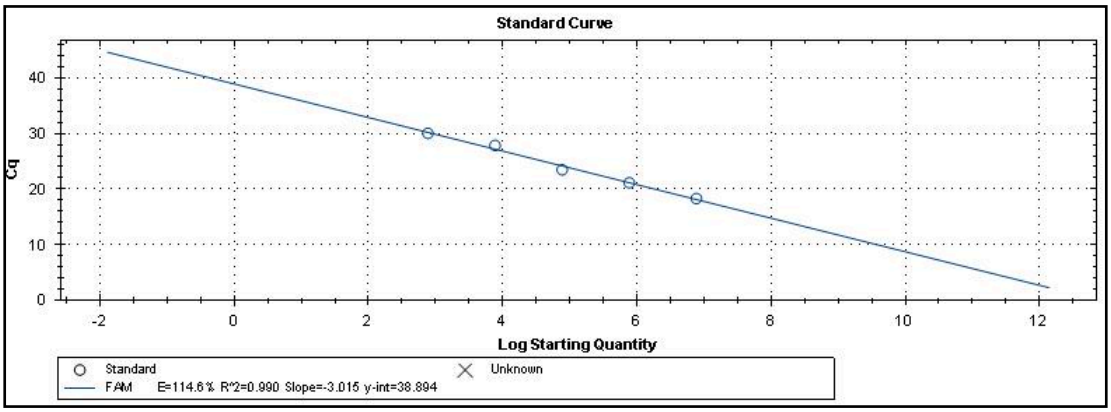
d-e-f. Sections of CF1 cells processed for transmission electron microscopy in 25000x direct magnification. The cells were processed and stained using 1% aqueous osmium tetroxide (OsO4) and uranyl acetate. The virions observed form large aggregates in the *extracellular* space (d) while budding virions were also observed (e-green tick mark). The diameters of observed virions are highly diverse (f) (range 74.2-97.1nm) whilst an immature-like particle structure is probable as significant density localised centrally in the capsular region is lacking in all virions identified. The budding virion shown (e) has a slightly more concentric structure at the capsule, while its diameter approximates that of a type C retrovirion at 86.6 nm.

4.2.12 Determination of the genomic content of CF1 cells in xenotropic MLV *gag*

To ultimately answer the question of whether the CF1 cells also bear non-activated; (ie not detectable via RNA PCRs) xenotropic or polytropic endogenous sequences, I directed already available XMRV/pMLV_*gag* primers (Tuke et al, 2011) versus gDNA from CF1 cells and various controls and cell line gDNA. The primers and probe have high specificity against xenotropic and polytropic MLV *gag* (Tuke et al, 2011). Human and mouse genomic copies control primers/probes (PDH and TFRC) were also used in a Real-time PCR format to allow for relative quantification of endogenous copies/ genome. A standard curve was constructed using ten-fold dilutions of the DNA plasmid Vp62 (Methods, 2.5.3). Genomic DNA was extracted using the DNAeasy Qiagen kit as described in Methods per manufacturer instructions. The results are shown in Figure 4.2.12.

Figure 4.2.12 Xenotropic/polytropic MLV *gag* real time PCR results on CF1 cells

A.



B.

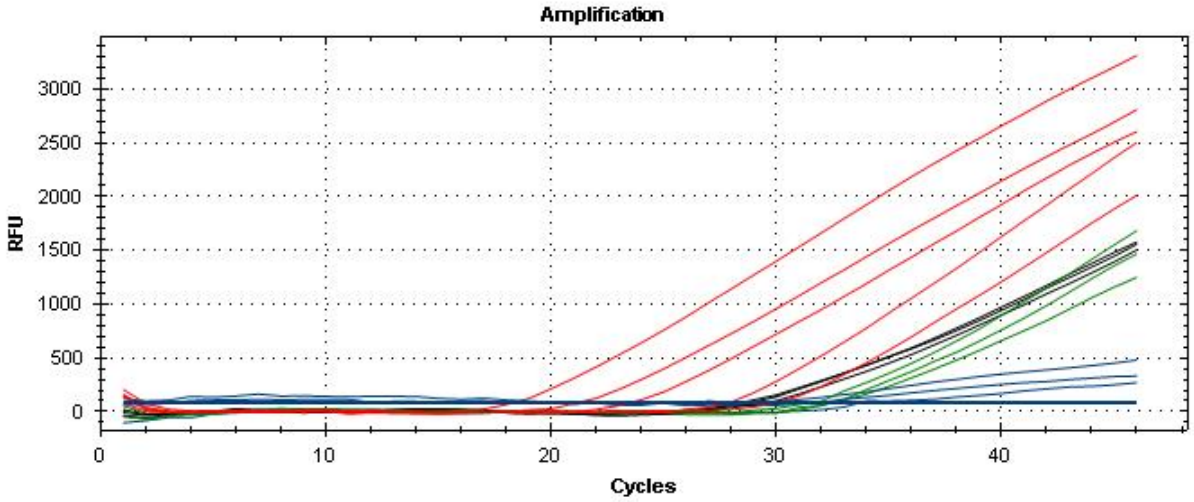


Figure 4.2.12 Xenotropic/polytropic MLV *gag* real time PCR results on CF1 cells

The use of XMLV/PMLV *gag* specific primers against genomic DNA from various cell line derived genomic DNA in Taq PCR. (A) shows the standard curve constructed using a serial ten-fold dilution curve of the VP62 (XMRV) DNA plasmid (R2:99% , slope: -3.015). (B) shows the standard curve amplification (red lines) and the amplification of various genomic DNA targets. The black amplification triplicate is human 293T cells infected with XMRV 6 days p.i. ; the green amplification triplicates are J2 cell line gDNA and the blue amplification is CF1 cell line derived gDNA. The reference mouse and human gene controls are not shown but amplified appropriately. The genomic content of CF1 cells in x/p MLV *gag* is measured as less than 1 copy per cell ie non amplifiable.

4.3 Discussion

In this chapter, molecular and imaging evidence has emerged of a consistent MLV RNA *gag* and *env* element present in CF1 cell-free supernatant,. This virus has aligned favourably to a consensus in direct sequencing of the envelope gene, which is specific for tropism, as an ecotropic MLV closely related to AKV, RFB or emv-30. As seen in the previous chapter this e-MLV is associated with a minor increase in RT activity. Interestingly active ecotropic *env* mRNA transcription is also experimentally demonstrated via transcription RT-PCR that produced quantifiable measurements and a sequencing confirmation.

Although TEM was able to visualise virion-like structures present in CF1 MEF cells, we observed that although C-Type, these virions do not exhibit a unanimous morphology neither appear to form a clearly distinguishable outer capsid layer that would also indicate replication competence. Transcription real time PCR however in the same total cell RNA demonstrate that the virus is actively replicating , splicing active mRNA for *env* and further the cells are positive in immunostaining for MLV-GAG in cytoplasmic and perinuclear regions of the CF1 cells.

Throughout our experiments, all individual sequences obtained from *gag* and *env* conserved primer targets aligned identically within replicates and samples from the same cells, growing throughout several time periods (Chapter 3) , indicating that there is one unique element present. However, the PCR amplification bias and selective bias at direct sequencing cannot be excluded. Further at analysis of the obtained sequences, the alignments never indicated a presence of a xenotropic variant, a hypothesis that was further established with the lack of a xenotropic/polytropic MLV *gag* signal in real time PCR. Further proof was sought via walking primer directed whole genome sequencing of the MLV virus from CF1 cells which however proved an impossible and lengthy technical task (data not shown). A further indication that the CF1-related RNA virus is of the non-xenotropic class, is also its inability to infect human DERSE cells that lack the mCAT-1 receptor for ecotropic virus entry, producing no immunofluorescence signal.

However, as the RFB MLV is reported to be infectious for mice and mouse cells, causing osteoma and osteopetroma in inoculated mice (Pedersen et al,1992) , we also assessed the CF1 MEF cell derived virus for infectivity on NIH-3T3 cells observing the appearance of characteristic infectious centres or syncytia. We observed that the CF1 cell- related ecotropic virus appears to infect the mouse cells as multinucleate syncytia are notable but does not propagate, with the number of syncytia being diminished with time.

The latter could be explained by the choice of our infectivity test cell line. NIH-3T3 cells are easily infected with ecotropic-MLV arising from common laboratory strains of mice, as they possess the entry receptor for e-MLV, m-CAT1 (Albritton et al., 1989). The NIH-3T3 cells traditionally produce 10-100 fold less cytopathic syncytia after an e-MLV infection than a rat cell line -the XC cells, does, and this is why the XC cells were historically commonly used for the quantification of e-MLV. Moreover, ecotropic MLVs are further classified into those that are able to replicate in NIH-3T3 cells (N-tropic), in Balb/c cells (B-tropic) and in both (NB-tropic) according to the host FV-1 restriction gene that has dominant alleles for resistance (Stoye, 2004). This restriction is post-entry and pre-integration and does not completely inhibit infection in vitro but rather reduces the number of infected cells by a factor of 50 to 1000 (Hartley, 1970).

Our CF1-MEF related MLV lack of replication in NIH-3T3 cells may hence be due to its B-tropism and a susceptibility to the NIH-3T3 produced Fv1n restriction factor, rather than a reduced infectivity. Notedly, the ecotropic RFB-MLV from CF1 cells is not able to productively infect the J2 3T3 cells, that are also N-tropic as shown by the lack of syncytia formation (data not shown).

The experiments described in this chapter appear to finalise our concerns over the polytropic *envelope* element in J2 cells indicated by sequencing and possibly associated with a minute RT activity found in Chapter 3. The J2 supernatant was indeed additionally used as a negative control for the MLV Gag Immunofluorescence herein where no MLV Gag expression was seen. The J2 cells also remained negative

after extensive analysis via TEM for virions (data not shown) and their polytropic content was seen further in the xenotropic/polytropic MLV *gag* RT-PCR, confirming our findings via direct sequencing. This does not preclude the likelihood of a recombination event occurring via superinfection. Their endogenous content is however unlikely to provide significant grounds for infectivity in standard conditions.

The use of CF1-MEF cells as stem cell culture feeders dictates prior mitotic arrest or inactivation that is achieved either with γ-irradiation or with Mitomycin C treatment. The first reports describing the derivation of hESC also described the necessity for the use of mitotically inactivated mouse embryonic fibroblasts (MEF) as feeder layers (Thomson et al., 1998 and Pera et al, 2000). It has been shown that a very high level of 100μg/ml and more is needed to restrict virus production when RNA synthesis is effectively blocked (Buck and Bather, 1968). This is in line with our finding that Mitomycin C at concentrations of 4μg and 10μg indeed does not influence virus expression, an experimental prerequisite for our further experiments with stem cell co-cultivations.

It is unclear why there have not been more reports of an expressed e-MLV virus from CF1 feeder cells, and this expression may prove problematical especially after the demonstration of a high amount of recombinant viruses mobilised by a replication-competent endogenous ecotropic MLV (emv-2) in mice *in vivo* (Young et al, 2012b). The implications of such an infection would be compromising, if not detrimental to the translational value of several findings that rely on the detection of endogenous retroviruses and/or viral integration of vector systems in human or murine stem cells and general molecular biology experiments. In a similar occasion, the inadvertent rescue of MLVs has been shown to confound viral cell culture results, as was the case with an env(-) HIV-1 rescue from Jurkat cells that was later found to be due to the expression of an endogenous MLV virus (Takeuchi et al, 2008b). An integration event, could be an absolute limiting factor in their applicability for human transplantation use. After attempts to characterise the CF1-cell related MLV virus with NGS, we will seek to answer this delicate question over the rare likelihood of a human cell

integration, taking advantage of the increased detection sensitivity available with this new molecular method.

Chapter 5
Full length sequencing of the CF1 -related MLV genome by RNA-Seq

5.1 Introduction

In this chapter I will describe the full length sequencing of the CF1 cell line-related MLV virus using a next generation sequencing (NGS) method (Illumina, MiSeq) based on RNA. Using NGS to generate a deep sequence allows for the detection of virus variants and a higher depth of analytical sensitivity. However, this is one of the few reports in the literature to date using viruses derived from cell culture supernatant and as such poses significant challenges which are outlined in detail in the following.

5.2 Methods

RNA extraction, library synthesis and NGS sequencing are described in detail in Methods. Input RNA was 0.1-1µg per library as assessed via spectrophotometric quantification. Two samples were prepared with equal concentrations of CF1-related MLV; one a pure CF1 cell supernatant and the other additionally spiked with carrier MS2 bacteriophage RNA, at a lesser concentration (8ng MS2 RNA per viral sample) that served as an internal control. We included two ultracentrifuged samples from 22rv1 cell supernatant containing $1x10^8$ copies/µl of XMRV *gag* as well; to control additionally for the sequencing efficiency of our protocol i.e. a positive virus control. Table 1 shows the input quantities as measured by the Agilent MicroChip BioAnalyser prior to sequencing and the total raw reads obtained.

Table 5.1 Samples submitted for RNA-Seq MiSeq NGS

sample submitted to RNA-Seq	mean fragment size bp	concentration (pg/ul)	total unpaired reads
CF1 sample a	294	809.89	4,798,261
CF1 sample b + MS2	344	8,959.91	4,964,654
XMRV sample a	321	1,334.22	1,529,905
XMRV sample b +MS2	294	28,587.37	824,748

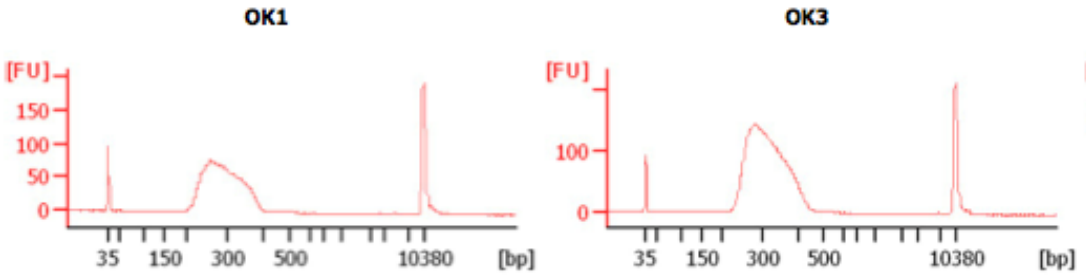
5.3 Results

5.3.1 Construction of libraries and assembly of full length viral RNA

The constructed libraries were quality assessed prior to submission for sequencing which was performed at the MRC Genomic Unit- Hammersmith Hospital, with a Bioanalyser RNA Chip for determination of fragment size, quantity and integrity. We first assessed the constructed libraries via gel electrophoresis (Figure 5.3.1A) and then gel purified the region containing the DNA segment below 150bp to remove primer dimers that lie in the region of 35-50bp. Figure 5.3.1B demonstrates the curves obtained for the samples submitted (CF1 +MS2 and XMRV+MS2) before and after gel purification.

Figure 5.3.1 Quality control electrophoresis of CF1 derived library pools

A.



B.

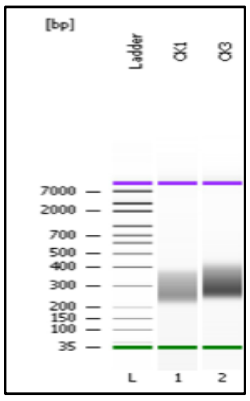


Figure 5.3.1 Quality control electrophoresis of CF1 derived library pools. Constructed cDNA libraries were analysed for size and quality using gel electrophoresis (A) and microchip electrophoresis (B). The libraries are consistent with cDNA fragments at the desired size of 300bp, while only a small peak is observed at the dimer length of 0-50 bp. Dimers formed at this size had been previously excised at gel purification, as they complicate further analysis.

To obtain the full length sequence of CF1cell line-related MLV RNA, I performed a series of assemblies of paired short length reads from next generation sequencing libraries obtained as described in Methods 2.5.10, as per manufacturer instructions. A combination of *de novo* and "mapping against reference sequences" from the GeneBank database to screen for the presence of known viral sequences, indicated previously from direct sequencing and the literature, was performed. The paired-end I7 indexed reads of the CF1 cell supernatant-related RNA libraries gave rise to an average 5 million reads for each duplicate CF1 supernatant sample and approximate 1 million reads for 22rv1 derived XMRV RNA (positive virus control) (see Figure 5.3.1A for numerical output of RNA-Seq). The MS2 bacteriophage internal RNA control was also efficiently amplified in both spiked samples.

Firstly, I assessed the number and quality of the internal MS2 control. In the CF1 sample 573,228 of 4,964,654 (11.5 %) aligned to the MS2 reference sequence at 20169X depth coverage, while for the sample containing XMRV, MS2 control carrier RNA was also efficiently assembled at 28,277 of 824,748 reads (3,24%) for a 32231X depth. In the positive virus control (XMRV full length RNA virus) 243,268 of total 1,529,905 reads mapped to the XMRV reference sequence (15.9%) whereas the rest of the reads mapped with ribosomal RNA targets (mouse and human), several bacterial species such as E.Coli and some identified as lentiviral cloning vectors which were assessed as likely contaminants from the sequencing platform. Ultimately, we were able to map the targeted paired-reads from 100% of the reference XMRV sequence for a depth of 3955X (range 1-9234, STDev ±1707). The contaminating sequences were found consistent in number and identity across all samples.

Of the approx. 5 million sequences from the first CF1 sample, 50,381 (1.05%) aligned to known murine retroviruses via the mapping-to-reference iterations. This amounts to the entire sequence-similar iterations as each murine MLV was mapped against the entire pool of reads each time individually. The individual percentages are shown in Table 5.3.1.1. A best-call base variant analysis with a stringent calculation for consensus support, using an in-house algorithm at the Bioinformatics Unit, Public Health England, Colindale that inputs quantification was therefore used. Depth of

coverage for the most prevalent and relevant alignments are shown in Figure 5.3.1.2. Finally, we tested our hypotheses with *de novo* assemblies from the entire pool that proved less detailed (Table 5.3.1.2).

Table 5.3.1.1 Mapping to reference sequence results for CF1 derived samples

Map to reference **assembly**						
Sequence ID	**Sequence name**	**% reads**	**Ref Length**	**Mean depth**	**%Ref-Seq**	**%Pairwise identity**
kj668270	emv-30	5.65254%	8448	94.2	100	99.1
kj668269	emv-30 (endogenous)	5.65303%	8867	83.3	99.95	99.6
dq366147	murine endogenous ecotropic MLV-1	5.79824%	8392	47.4	99.9	99.1
u43202	RFB-MLV	6.39017%	9045	44	99.9	99.1
j01998	AKV	5.65012%	8382	38.6	98.4	99.5
ak090127	murine endogenous MLV eMLV-2	5.73503%	8070	94.2	98.1	99.1
mlu13766	MCF virus	3.25073%	8232	57.6	82	99.5
kj668271	polytropic MLV NOD/ShiIJ	9.08233%	8209	27.5	79.3	99.5
xmv43	xenotropic endogenous MLV-43	3.73642%	8802	35.4	73.9	99.5
v01184	FBJ-MSV	0.882408%	4236	16.3	57.6	99
af033814	murine osteosarcoma virus	1.19181%	3878	12.6	57.3	99.4
MUSIAPL31	IAP element clone L31	Not found	3889	0.0	3.3	97.7
MUSIAPP3A	IAP element clone p3A67	Not found	2213	0.0	0.0	0.0

Table 5.3.1 Mapping to reference sequence results for CF1 derived samples

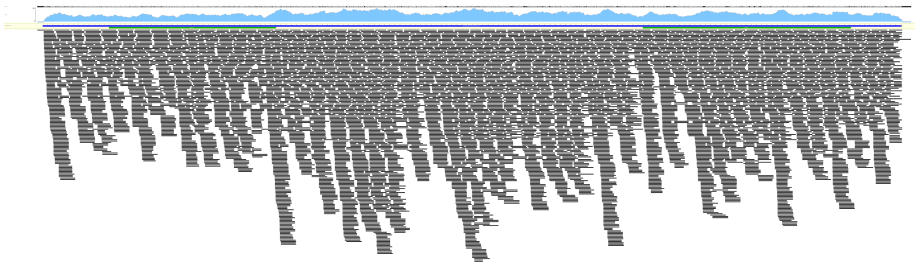
Map to reference assemblies were performed in the reads arising from the sample containing CF1 cell supernatant virus and MS2 positive RNA carrier control. The table shows the sequence IDs, the name of the viral isolate, the percentage of the isolate present in the total reads (%reads), the length of the reference sequence (Ref length), the number of reads on average for the assembly (depth) and the percentage of the RefSeq covered (%RefSeq). Last, the % pairwise identity of the alignment between the assembled consensus and the reference sequence is shown as % pairwise identity.

The best reference sequence match (Table 5.3.1) for CF1 samples was that with emv-30, an endogenous ecotropic MLV from NOD/ShiLtJ mice determined at 100% coverage of 8245bp with a mean depth of 94X (Range 1-167, STDev ±27.2) and 99.1% pairwise identity (Figure 5.3.2A). With lesser scores (98.4% of RefSeq, depth 38.5X ±14.8) a match was also obtained for the AKR endogenous MLV (Acc No j01998). The mapped assembly however had a significant gap in the U3 repeat region LTR (bp 7968-8060) that is characteristic of AKV (data not shown). This is indicative proof that in the cDNA library tested, there is no sequences that match for the AKV full virus.

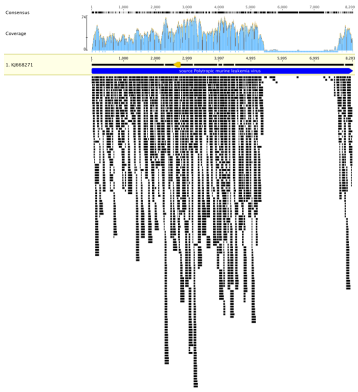
The RFB-MLV full length RNA was also not supported by the full length assemblies, as the mapped reads only covered 95.5% of the RFB-MLV (7,982 of 8,357bp) with a gap in the p-15 region of *env* (7345-7566 bp). This region had not been previously covered by our *envelope* direct sequencing primers that cannot discriminate between emv-30 and RFB envelope and are conserved for all MLV tropism (See Chapter 4). FBR-MSV was also not found as a large portion of the viral sequence length (57.2% of reference sequence, depth 16X) was not found in the samples, with a large gap in the region of the p55 oncogene (Figure 5.3.2C). A significant finding is that the polytropic equivalents were also not present with a gap at the *envelope* gene that failed to assemble (0 sequences) (Figure 5.3.2B). Xenotropic reference mapping also produced largely deficient alignments in the *gag* and envelope regions (data not shown).

Figure 5.3.1.2 Assembly results from paired-end reads of next generation RNA sequencing on CF1 derived pools

A.



B.



C.



Figure 5.3.1.2 Assembly results from paired-end reads of next generation RNA sequencing on CF1 derived pools

A**. CF1 read pool assembled against emv- 30 (Acc. kj668269).** 5960 paired end reads (black bars) were assembled against the 100% RefSeq of emv-30 for the full length of 8448 bases. The coverage depth for the alignment is on average 94.2±27.2 (STDEV) (Range 1-167) which is significantly high. Coverage average variation is shown within the blue graph.

B. CF1 read pool assembled against polytropic MLV-30 (Acc. kj668271). 3020 paired end reads (black bars) were assembled against the 79.3% RefSeq of polytropic MLV-30 for a partial length of the 8209 bases. The gap observed is at the region bearing the right end of the *env* gene, which is different for polytropic viruses.

C. CF1 read pool assembled against FBR-MSV (Acc. V01184). 522 sequences were assembled against the 57,6% RefSeq of FBR-MSV for a partial length of the 4236 bases. The gap observed is at the region bearing the FBR_MSV associated *v-fos* oncogene.

We then performed a two-tailed algorithm where obtained sequences were assembled *de novo* using Geneious Assembler and mapped to reference MLV sequences from the Genbank database (Acc No AF033814, AK090127, DQ366147, DQ366148, FN692043, J01998, KJ668269, KJ668270, KJ668271, NC001417, U43202, V01184). *De novo* assembled contigs were blasted against the nr/ns nucleotide database. (Table 5.3.2). A contig sequence of maximum 2549bp was constructed via *de novo* assembly that mapped via Blastnr to AKV (Acc No J01198) and emv-30 (Acc No KJ668269) equally at 99.2% pairwise identity, at the *gag* region (RT-IN) with a gap in the start of the *Env-gp70*. A second contig was mapped via Blastnr to mouse endogenous MLV emv-3 (99.5%) and AKV (98.4%) containing the right but not the left half of *Env-gp70* and the full *Env-p15*. Again, the alignment with RFB suffered significantly at the region of the *p15* (7345-7566bp).

However, although the *Env-gp70* was partially missing from the *de novo* sequence, it was fully present in the library pool as shown by mapping to reference assemblies. This discrepancy is further discussed in Section 5.6 associated with the *de novo* assembly limitations. The same is not the case for the RFB-*p15* that has a number of differences at the nucleotide level to the RNA sequence of the *p15* present in the tested CF1 cells. Table 5.2A provides with a detailed analysis of the *de novo* assembled contigs in CF1 cells, including the content of the supernatant in sequences other than MLV viruses for the purposes of quality control.

This further strengthened the hypothesis that the virus present in CF1 cells is closer consensus to emv-30, as this was the least variable and mapped fully 100% and further analysis was performed based on the full length sequence obtained from this emv-30 mapping.

Table 5.3.1.2 *De novo* assembly results from CF1-derived samples in RNA-Seq

***De novo* assembly**			
272,990 of 671,756 reads assembled *de novo* to produce 6,115 contigs			
10 contigs >1000bp			
Reads assembled	**Sequence length**	**Organism**	**% Pairwise identity**
149	2.549	AKR(endogenous) murine leukemia virus	99.2
67	1.418		98.4
35367	1370	murine rRNA	99
20371	1334	murine mRNA	91.8
20300	1043	murine mRNA	93.1
4257	1547	murine rRNA	94.7
3877	3585	Enterobacteriophage	99.9
2234	1267	murine rRNA	95.4
791	1047	murine rRNA	95.3
271	1110	murine rRNA	95.4

Table 5.3.1.2 *De novo* assembly results from CF1-derived samples in RNA- Seq

De novo assembly produced 10 contigs above 1000bp from a total of 671,756 paired-end reads in the sample containing CF1 cell supernatant virus and MS2 positive RNA carrier control.

5.3.2 Detailed organisation of the CF1 cell line-related full length consensus sequence

The viral RNA obtained from the RefSeq mapping of 100% identity is 8,448bp in length (the full fasta sequence GenBank Accession MG525170 is available in Appendix 6). The R repeat sequence of 68bp length is present in the 5'LTR preceding the U5 region, and also in the 3'LTR in the region between the U3 and U5. The U5 sequence is 74bp and is followed by the 18bp of the primer binding site (PBS) coding for tRNAPro and the 5' untranslated region in *gag* that encodes for the packaging Ψ signal. Contained within it, is the CUG codon that initiates glyco-Gag; the "normal" Gag protein is coded starting at position 627 where the AUG codon is found. The *gag* and *pol* coding sequences are in the same translation frame; they are separated by a single UAG termination codon.

The messenger RNA *env* splicing donor sequence is present at position 522 in the 5'LTR. The splice acceptor sequence is observed at the beginning of *env*, within the 68bp overlap with the end of *gag* at position 5787bp and is identical to the splice acceptor for wild type AKV virus. No alternative splice donor or acceptor sites are identified, as the ones described for the AKV virus (Sørensen et al. 2007) that may lead to the expression of trapped *gag* sequences. The polypurine tract (PPT, 7800–7822) is followed by the U3 (381bp length) and R region.

No abnormal termination codons were found within the paired-end read consensus, but two variants were discovered causing nonsense substitutions in 3% (A4346G within RT) and 8% (A5927G within IN) of the assembled reads (mean coverage in the regions, 55X) (further analysed in the subsequent section). Moreover, in the assessment of insertions or deletions we identified two one-nucleotide deletions a) in position X784C, that falls within the start of the polyGag coding sequence and is represented by 9% of the reads with 31X regional coverage and b) in position X7753T, that falls within the transmembrane envelope protein (TM) region and is represented by 6% of the reads with 38X coverage. The importance of these deletions to coding potential and protein functionality are analysed in the following section. Insertions

respective to the reference sequence were observed in less than 1.3% of base reads in two positions and are not considered.

5.3.2.1 Coding regions and open reading frames

As in all orthoretroviruses, the organisation of the coding regions are from 5' to 3', *Gag:Pol:Env*. Traditionally, the *Pol* proteins are initially synthesised together with *Gag* in a large *Gag-Pol* polyprotein from unspliced full length transcripts. The CF1 MLV virus has both coding regions, separated by a single termination codon at position 2241.

This codon is subject to stop codon read through at the ribosomal sites, where some 5-10% of translation bypasses the termination signal and goes forward inserting glutamine to produce the full *Gag-Pol* polyprotein. This recoding process was until recently thought to be primarily directed by a 57-base cis-acting element immediately 3' of the termination codon (Felsenstein&Goff 1988); which is indeed present in the CF1 MLV. This bipartite signal induces the insertion of glutamine in 5% of the translation products; a process sought to be achieved via steric hindrance of ribosome function due to the presence of an active RNA pseudo knot in the region (Alam et al, 1999). There is no evidence to suggest that the full length MLV in CF1 cells is translating *GagPol* in a two frame specific pattern, as functional protein is translated in the same frame (1 forward) in the sequence obtained, and no alternative splice donor or acceptor sites are identified.

For orthoretroviruses, the *Gag* polyprotein is cleaved at protease sites to produce the full proteins of the Matrix protein (MA) or *p15*, Capsid (CA) or *p30*, Nucleocapsid (NC) or *p10*. The *Pol* protein is cleaved to produce the full proteins of Reverse Transcriptase/Ribonuclease H (RT) and Integrase (IN). Cleavage occurs during or shortly after the release of the virion from the plasma membrane in a cascade-like process that is part of the virion maturation. No termination codons to indicate defective gene function are found.

Mature proteins are relaxed by proteolytic cleavage by the retroviral protease (*p14*), with the exception of the integrase domain that is separated from the *Pol* polyprotein by action of its own protease. The retroviral protease or *p14* for CF1 MLV is found to exhibit aspartyl protease activity.

The *envelope* gene is translated in a second frame in MLV viruses, from a spliced transcript. This originates from a splice donor site (SD) at the 5'LTR, present in position 213 and a splice acceptor (SA) at the beginning of *env* at position 5774. The envelope sequence comprises of the surface protein (SU) (*gp70*) and the transmembrane protein (TM) (*p15E*) that are produced associated and form trimers. The receptor binding domain is located in the surface subunit, with three distinct regions, VRA, VRB and VRC fully recognised. The TM subunit contains a fusion peptide of 20aa, that is characteristic for its heptad repeats; i.e. two leucine zipper like regions that are capable of forming into a coiled-coil structure.

No further accessory or oncogenes were recognised in the sequences at the central region or the 3' end of the CF1 MLV RNA viral genome.

5.3.2.2. Upstream non-coding regions

The proviral sequences at the 5' end of the retrovirus, i.e. preceding the first open reading frame, are the 5' long terminal repeat and the 5' untranslated region. Both have significant numbers and variety of regulatory elements, enhancers and transcription factors, that are necessary for adequate interaction of proteins and nucleic acids.

5'LTR : The size of the CF1-MLV LTR is 524 bp long. The U3 region covers most of the 5'-LTR, demonstrates the attachment sites (att) that are necessary for the action of integrase and the 3'end processing for integration. (Chow et al,1992). These are imperfect inverted repeats that have TG and CA as highly conserved nucleotides. The U3 region in the 5'-LTR serves as a promoter by driving cellular polymerase II transcription. The core promoter region contains a TATA box, traditionally bound by the multiunit TFIIB complex.

Regarding the non-coding regions, that are essential for viral fitness, replication and integration, the full 18bp tRNAPro PBS at position 445, is found located ~136 bases from the 5′ end of the RNA and ~460 bases 5′ from the start of the *Gag* coding region. The Polypurine tract (PPT) is found at the 3' end of the RNA at pos 8113 and is 23 bp long. The polyadenylation signal is located in the R region of the LTR and the U5 region. This is a distinct characteristic present in MLVs and means that the 5'poly adenylation signal codes for weak poly(A) to facilitate genomic RNA formation.

The ψ packaging signal, located normally at the non-coding region between the PBS and the start of *Gag* (leader region) consists of intermolecular stem loops, that for MLV bear CACG in the loop. MLV-RNA also contains two motifs with the sequence UCUG-UPu-UCUG that are a crucial element for the dimer packaging of viral RNA. CF1-MLV contains two CACG loops at pos 552 and 639 and the 4 UCUG component motifs within the psi region, which indicates that it is competent to package dimeric RNA.

5.4 Variant determination

For the firm establishment of a consensus derived from the numerous paired-end reads of the sequencing output, it is recommended that a detailed analysis of the variants present in the analysed sample is performed. This can be aided with use of available statistics generated by the analysis software during the construction of the consensus. The analysis between all variants present in >10% of the reads against the consensus CF1-MLV, obtained through the alignment, are shown in Table 5.4. Variants with a percentage representation above 10% of paired-end reads and adequate depth coverage (>30X) were considered significant. An additional analysis was performed using SNAP2 (See Methods 2.5.10) to distinguish silent mutations, amino acids (AA) substitutions, stop codons and gaps/frameshift mutations amongst variants. SNAP2 as well as SIFT were also used to identify mutation function effect at the protein level.

Table 5.4 Variant base pair determination in the CF1-MLV consensus

Variant calling										
region	**Silent**	**%**	**cover**	**region**	**AA substitutions**	**Mutation**	**%**	**cover**	**SNAP2 (%)**	**SIFT**
Gag	T466C	26	38	*gag*	C521T	P513S	18.6	59	neutral 78%	tolerated
	G626A	25	34	*gag*	G1008A	A618T	19.5	34	non-neutral 63%	tolerated
Gag-MA1	C658T	25	31	*RNAse H*	A4532G	R761G	11	54	non-neutral 87%	tolerated
	G2215A	36	44	*Integrase*	A4610T	T14S	23	43	neutral 69%	tolerated
	T2359C	14	61	*env*	A7550G	E463K	10	72	non-neutral 63%	tolerated
	G2371A	13	56	**region**	**Stop codons**					
	G2389T	13	43		A5927G		8	52		
	T2392C	12	49	**region**	**Insertions**					
	G2767A	17	72		insT3456		1.3	72		
	A3493G	27	41		insT4864		1.3	78		
	C4714T	13	62	**region**	**Gaps/frameshift**					
Integrase	A4828G	16.6	54		X784C		9	31		
env	G5119A	10	59		X7753T		6	38		
	G5569A	10.5	38							
	G6921A	15	42							
Non-coding										
	G8128A	29	45							
	A8209G	14	53							

Table 5.4 Variant base pair determination in the CF1-MLV consensus

Base pair differences in the paired-end reads of at least 10% in the obtained sample, are presented along with an analysis of their predicted effect on protein function using two online protein prediction software platforms (SNAP2 and SIFT). Only 5 are discovered to cause amino acid substitutions, of which none has a detrimental effect on function, although SNAP2 analysis identified three as non-neutral i.e. likely to have some effect. The stop and gaps mutations are by definition detrimental to protein function.

5.5 Alignment of the full-length CF1 MLV consensus

The consensus sequence for CF1 MLV obtained via the best reference mapping assembly was aligned with several related ecotropic MLV sequences, using CLUSTALW. The closest similar sequences i.e. emv-30 and the AKV endogenous virus (MLOCG) are shown in Figure 5.5, in comparison to the CF1 consensus. The main nucleotide substitutions are marked for each compared sequence. Thirteen single base pairs differ in the alignment between CF1 MLV and emv-30 (KJ668269), of which only three (A2001C-A6267G and C7774T) are amino acid substitutions and subsequently of which only one (A2001C) codes for an intolerant protein mutation (*gag* N415H), i.e. predicted to have effect on protein function. This is due to the substitution in the coding triplet from histidine (H) in emv-30 (positive aromatic) to asparagine (N) (negative and non-aromatic).

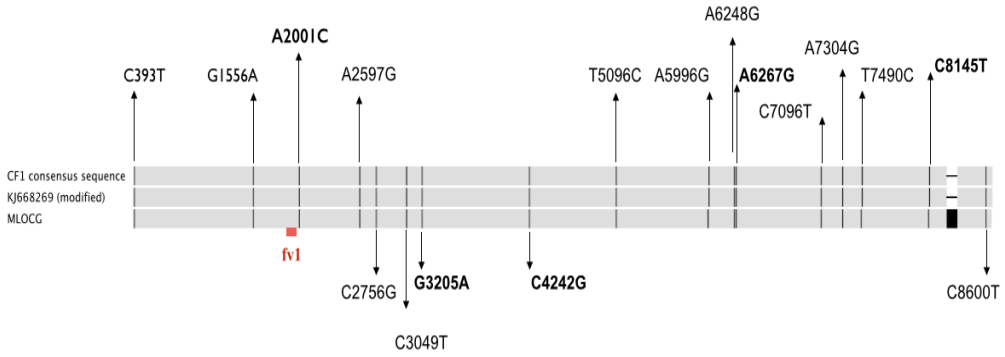
Interestingly, the N415H mutation is located in the region coding for the Capsid protein *Ca* and in particular, the MHR (major histocompatibility region) as shown in Figure 5.5. Several authors have described mutations in MLV isolates in the particular area (Kozak et al, 2006; Mortuza et al,2008; Ohkura et al, 2011) that are hence defective in budding and infectivity, however this CF1 virus mutation appears not to have been previously annotated. The mutation is 100% supported in the consensus sequence i.e. there is no variation present for the particular base in the total reads pool. Further information is presented in the discussion section of this chapter.

We also compared the CF1 consensus sequence to the endogenous MLV emv-2 that is defective in nature due to a mutation in the *pol* gene and the RTase. Forty (40) base pair differences are discovered with emv-2 (not shown) but amongst them the recently described *pol* inactivating mutation that renders emv-2 defective is not seen. In contrast, the CF1 consensus appears to have a functional RT-coding gene region. An extra five (5) to the 13 base pair differences between CF1 MLV and emv-30 are found between CF1 MLV and AKV-MLV (MLOCG). These are three additional amino acid substitutions at C3049T coding for A139V (tolerated), at G3025A coding for R191K

(tolerated) and C4252G coding for T540R (tolerated). One is a silent mutation at C2756G and one is the known 99 bp insertion of AKV-MLV at position A8394.

Figure 5.5 Mutation analysis of CF1 cell-related MLV

A.



B.

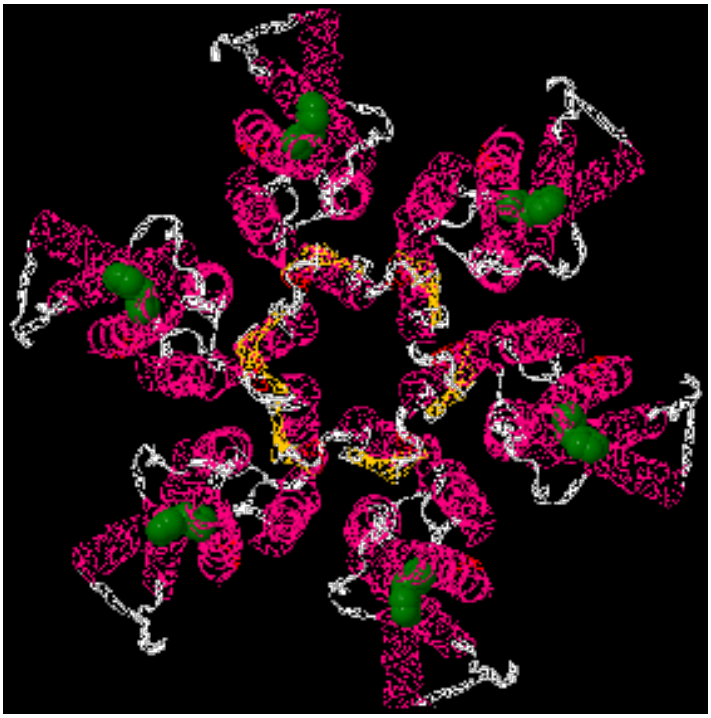


Figure 5.5 Mutation analysis of CF1 MLV against related MLV isolates

A. **The main differing base pair characteristics of the CF1 MLV and its closest phylogenetic counterparts emv-30 (KJ668269) and AKV-MLV (MLOCG) are aligned.** Arrows indicate the location of base pair substitutions between compared isolates. Arrows pointing on top refer to differences between CF1 MLV consensus and the KJ668269 emv-30, while arrows pointing down refer to the additional nucleotide differences between MLOCG AKV-MLV and the CF1 consensus. Note that in bold are mutations that lead to AA substitutions. In red is depicted the location of the restriction factor fv1 coding region.

B. **The AKV N-terminal Capsid Hexamer Structure (Uniprot P03336- PDBe 1u7k) and the N328H CF1 MLV related protein-altering mutation.** The Capsid hexamer structure (ie lattice form) for AKV-MLV was discovered in 2004 using high resolution X-ray diffraction at 1.85Å (Angstrom) by Mortuza et al. The sequence is identical for CF1 MLV apart from one nucleotide change that leads to an AA substitution in the region depicted as green areas, where Histidine (H) is replaced with Asparagine (N). The main function of the Capsid hexamer is during virion assembly. Further discussion is provided in the Discussion section of this chapter below.

5.6 Discussion

In this chapter the use of next generation sequencing to obtain the full length sequence of an unknown MLV retrovirus proved successful. It is true that this comprised of a semi-automated mechanism, while several researchers are now able to fully automate with the development of novel computer networks (reviewed in Sharma et al, 2015) . With this assay however, we have been able to exclude the presence of several of the query RNA viruses that had proved unable to differentiate via traditional direct sequencing, namely the polytropic and xenotropic MLVs and any oncogenic counterparts in growing untreated CF1 cells. By using a positive cell culture derived virus control as well as a positive carrier RNA control we validated our protocol for the characterisation of viral RNA from tissue culture supernatant and achieved sufficient depth of coverage for the determination of variants.

The full length characterisation of CF1-related MLV shows that beyond doubt the described viruses from CF1 mice in Pedersen et al, 1992 (RFB, FBR or FBJ) are not being expressed fully by the same name CF1 cells that are used in biomedical research as MEFs. The branches of cross breeding and distribution (as shown by original research in the Introduction Chapter) are long and intertwined. The description of the full length virus in the CF1 cells obtained in this thesis from ATCC SCRC-1040, indicates a close phylogenetic relationship with AKV-MLV that is traditionally thought to be present only in Japanese origin *mus mollosinus* mice breeds. It is also much closer related to emv-30, a described ecotropic MLV (Prochazska et al, 1991) also found in NOD/SHiLtJ mice (Triviai et al, 2014). It is therefore likely that CF1 have a common parent with NOD mice.

It is interesting that the CF1-related MLV sequence sheds some additional light over the functions that the virus is shown to demonstrate in Chapter 3 and 4. In this aspect, the CF1 related virus is not shown to exhibit high RT activity; however, there appears to be no functional defect in the *pol* gene (as in the emv-2 described in Young et al, 2012a). Moreover, CF1-related virions, as shown by TEM are irregular shaped and appear defective and/or immature also unable to induce cytopathic syncytia on mouse

cells. The defect in the Capsid polymer protein as shown by the N415H mutation, that is located in the region coding for the Capsid Protein Ca and in particular, the MHR (major histocompatibility region), maybe responsible for such a virion morphology; as no other structures were found affected by the coding sequence in the CF1 cell supernatants. In contrast, the sequence rather suggests of a competent ecotropic MLV sequence that is different from emv-30 and AKV ie distinct but closely related.

This sequence therefore, in combination with the results from viral infectivity assessment and molecular experiments indicates that there is very low risk for human stem cells co-cultivated with CF1 cells, the power of NGS to describe even at the unique cell level, retroviral integration events is a necessary exercise for the assessment of the safety of sensitive human material used in clinical transplantation.

Chapter 6
Viral integration analysis

6.1 Introduction

In this chapter, I used splinkerette PCR followed by direct Sanger sequencing and NGS to analyse MLV DNA integration into various mouse and human cell lines, after co-cultivation.

6.2 Integration site analysis via splinkerette PCR for CF1-related MLV in genomic DNA

6.2.1. Retroviral integration in genomic DNA from cell lines

A protocol that is specifically adapted for the determination of the integration sites of retroviruses in genomic DNA- the splinkerette PCR, was initially used for the discovery of the integration site of the CF1-related MLV in CF1 mouse cell DNA (See Methods 2.5.9). A positive control cell line was YAC-I cells (ATCC TIB-160) that is *Mus musculus*, strain A/Sn, lymphoma tissue infected with Moloney MLV -a prototypical MLV that is lymphomagenic *in vivo*. The YAC-I gDNA was anticipated to bear several integration sites of the Mo-MLV provirus, making it a good positive control for the splinkerette reaction from the digestion (step 2) and onwards up to the amplification stage.

Genomic DNA was extracted from YAC-I cells (ATCC TIB-160 100 ng /µl gDNA). The primers used were the Mo-MLV specific LTR primers found in Adams et al, 2009. At least 8 bands are observed in the integration PCR product from YAC-I cells with the splinkerette method (Figure 6.2.1.1).

Figure 6.2.1.1 Positive MoMLV virus integration control YAC-I

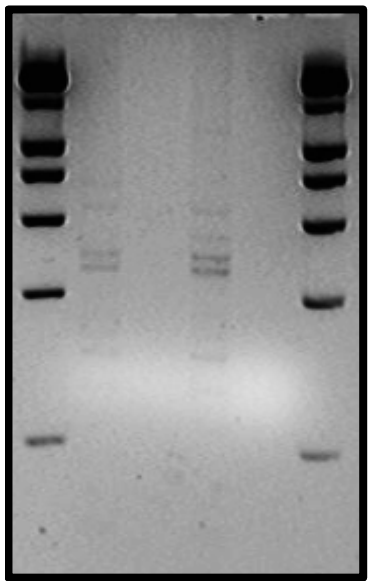


Figure 6.2.1.1 Positive MoMLV virus integration control YAC-I (Murine lymphoma mouse cell line), was used in the splinkerette reaction as it bears several integration sites from MoMLV(Moloney MLV). Lanes 1-3: YAC-1 genomic DNA. Lanes 2-4 Water negative controls.

I proceeded to the adaptation of the published nested primers, for the XMRV viral LTR (5'-end) and specifically the U3 region. The primers were tested against 22rv1 cell gDNA; previously at least 10 integrated copies per 22rv1 cell have been shown with southern hybridisation in Knouf et al, 2009. The primer sequences are found in Appendix. 1.

The single positive amplification band in 22rv1 cells is shown in Figure 6.2.1.2. The direct sequencing product mapped two identical PCR products that included the adaptor, the viral LTR and a human genomic sequence that was subsequently mapped in Blast, as a unique unambiguous hit in human chromosome 2 (RP11-617F9, AC022210.8, position 75,126, upstream). The result was mapped against the human reference genome at Chr 2; 85,040,544 in sense orientation (Acc No AC022210.8 cytoband p11.2). The integration is 54K bp downstream of the Thymosin Beta gene (TMSB10) and 11.2KBp upstream of an opposite strand Pottasium channel coding gene (KCMF1). At least four LINE elements were discovered between 84bp and 2695bp of the insertion site that lies 11.3KBp / 54.3KBp away from CpG islands. The sequence can be seen in Figure 6.2.1.3.

Figure 6.2.1.2 XMRV integration site in 22rv1 gDNA

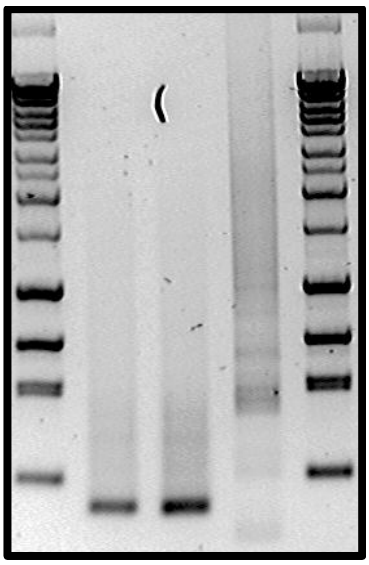


Figure 6.2.1.2 22rv1 cell gDNA by the XMRV virus (lanes 1-2), the YAC-I positive control amplification (lane 3). The two bands were sequenced and the integration site was mapped to Human Chromosome 2; 85,040,544.

Figure 6.2.1.3 The integration sequence obtained via direct sequencing for XMRV in 22rv1 gDNA

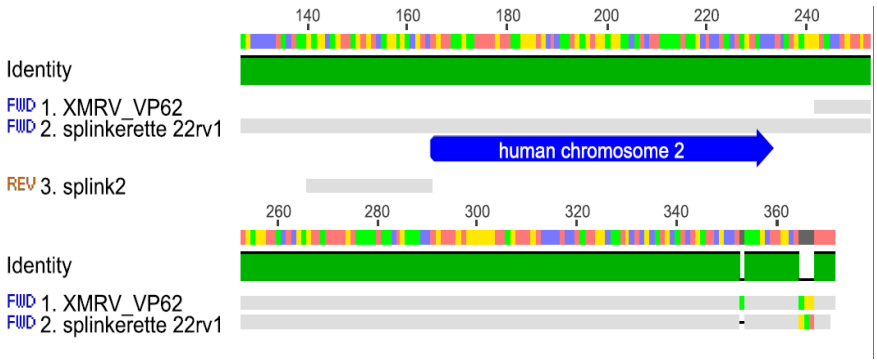


Figure 6.2.1.3 The integration sequence obtained via direct sequencing for XMRV in 22rv1 gDNA The sequence obtained from the splinkerette reaction in 22rv1 cell line gDNA is shown aligned with the XMRV_Vp62 3'end LTR and the splinkerette primer-adaptor (splink2) at 100% alignment (grey bars). Human chromosome ch2 at position 85,040,544 gDNA is annotated in blue.

A new set of nested CF1-MLV specific primers that are conserved for several ecotropic MLVs at the U3 region of the 5'LTR were also designed (Appendix 1). Although the reports concerning the CF1 related viruses are limited, a recent publication has described an endogenous retrovirus insertion site in CF1 feeder cells, when these were used in murine induced pluripotent stem cell propagation (Quinlan et al, 2011). That publication mapped the new integration to Chr 8; 125,949,348-125,958,700 (*Mus musculus* mm9). The result of our splinkerette amplification is shown in Figure 6.2.1.4. One amplification well failed to accurately amplify and appears undigested (Lane 2). The successful result mapped to Chr8; 148,187-148,317 (*Mus musculus* mm9) and the alignment of the PCR product is shown in Figure 6.2.1.5. We therefore enquired the generic mouse genomic sequence (mm9) for chromosome 8 at the position indicated from direct sequencing at position 148,187-148,317; we obtained a full length viral MLV sequence that is not RFB but better aligns with pmv10 and xmv12 (Jern et al, 2007). We have therefore now discarded this result as the true integration site of the *expressed* CF1-related MLV as neither polytropic nor xenotropic sequences were demonstrated in our CF1 cell supernatants (Chapter 4 and 5). This however does not preclude the likelihood that such a copy is present, albeit not expressed in the induction experiments prior detection (Chapter 3-4) or the stem cell co-cultivations (see below in Chapter 7). Unfortunately the CF1 cell genome has not been published for comparison.

Figure 6.2.1.4 CF1-related MLV integration site in CF1 gDNA

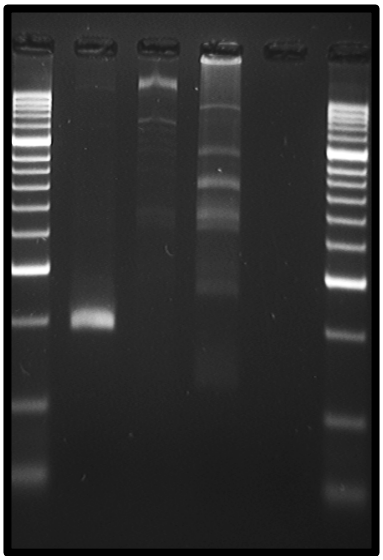


Figure 6.2.1.4 CF1-related MLV integration site in CF1 gDNA In the gel electrophoresis above is shown: CF1 cell line gDNA (lane 1), a failed amplification replicate (lane 2), the YAC-I positive control amplification (lane 3) and a water negative control (lane 4). The two amplicons were sequenced and the integration site was mapped to Mouse Chromosome 8 (see Figure 6.2.1.5).

Figure 6.2.1.5 The integration sequence obtained via direct sequencing for CF1-related MLV in CF1 gDNA



Figure 6.2.1.5 The integration sequence obtained via direct sequencing for CF1-related MLV in CF1 gDNA. The sequence obtained from the splinkerette reaction in CF1 cell line gDNA is shown aligned with the RFB-MLV 5'end LTR and the CF1 splinkerette primer (one mismatch). Mouse chromosome ch8 at position 148,187-148,317 (mm9) gDNA is annotated in purple. A portion of the adaptor is found in the beginning of the sequence-not annotated.

6.2.2. Retroviral integration in genomic DNA post *in vitro* infection

The developed integration PCR was assayed on infected cells gDNA to determine the infectivity of the CF1 related endogenous MLV for other mouse feeders and human cell lines. I experimentally infected growing MEF cells (J2 origin and NIH-3T3 origin) and human 293T cells *in vitro*, by inoculating CF1 cell supernatant (1ml in T75 flask of growing cells; estimated to contain 10^6 MLV RNA copies). At 7dpi I extracted the gDNA of the infected cells and the uninfected controls with the Qiamp DNAeasy extraction columns (Qiagen, UK) and performed the integration PCR, the results of which are shown in Figure 6.2.2A and B.

Figure 6.2.2 CF1-related MLV integration in infected mouse cell line gDNA

A. B.

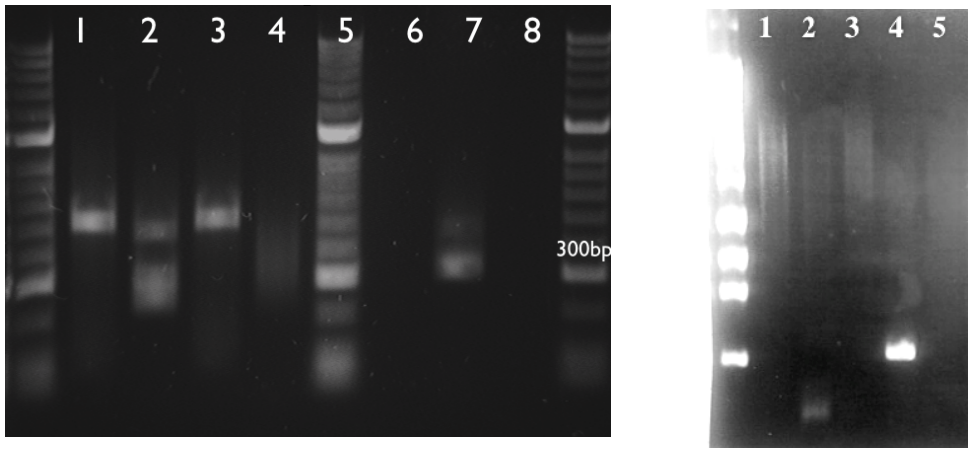


Figure 6.2.2 CF1-related MLV integration in other mouse and human cell lines

A. **J2- 3T3 gDNA from cells infected with CF1-related MLV**: Lane 1: uninfected J2 gDNA Lane 2: J2-3T3 gDNA infected with CF1-MLV Lane 3: 293T human gDNA uninfected Lane 4: 293T gDNA infected with CF1-MLV Lane 5: molecular ladder 50bp Lane7: CF1 gDNA positive control. Lanes 6 and 8: Water negative control.

B. **NIH-3T3 gDNA from cells infected with CF1-MLV** Lane 1: uninfected NIH-3T3 gDNA; Lanes 2-3: CF1-MLV infected NIH-3T3 gDNA in duplicate; Lane 4: CF1 cell gDNA positive control; Lane 5: Water negative control.

The result of the CF1-MLV infected J2-MEF amplification is shown in Figure 6.2.2A lanes 1 (uninfected) and 2 (infected J2 cell gDNA). Interestingly, we also included the experimental *in vitro* infection results from human 293T cells, uninfected (Lane 3) and infected with CF1-MLV (Lane 4), two water negative controls (Lane 6 and 8) and our CF1 gDNA positive control (Lane 7). Lane 1 integration in uninfected J2 cells maps to murine chromosome 14 (position 67,691,414). Lane 2 from infected J2 cells failed to produce a valid result both for the top and the bottom band after gel excision and purification. Lane 3 from uninfected 293T cells also mapped to mouse chromosome 14 (NT_039606, position 67,691,408).

Sequencing of the band in Figure 6.2.2B- Lane 2 initially indicated that the NIH3T3-MEF cells after infection with CF1-related MLV bore one new proviral integration, confirmed by the amplification of the MLV-specific LTR and an insertion sequence in mouse genomic DNA, that is absent in uninfected MEFs (Lane 1). The CF1 gDNA aligned comparably to chromosome 8 as per the previous splinkerette assays (positive control) (Figure 6.2.2A). The NIH3T3-MEF gDNA insertion sequence was mapped to mouse chromosome 14, position 67,691,392 (Acc No NT_039606) cytoband D1.

Interestingly, all gDNA were previously assayed with human (PDH) and mouse (TFRC) reference RT-PCR prior to the splinkerette assay, and 293T cell gDNA had amplified positive for the human PDH gene and negative for the mouse genomic control TFRC (Figure 6.2.3A). An alignment of the contaminant sequences is shown in Figure 6.2.3D. The effect of mouse DNA containing columns has been discussed previously and was the reason why mouse free DNA extraction columns had substituted our previous columns. Indeed, a splinkerette experimental assay with the uncertified mouse DNA containing columns showed extensive contaminating insertion amplicons, throughout infected and uninfected samples as well as water controls (not shown).

Interrogating the Ensembl Reference genomic sequence of mouse chromosome 14 at the region of the discovered insertion site, with the use of Retrotector, a tool used to

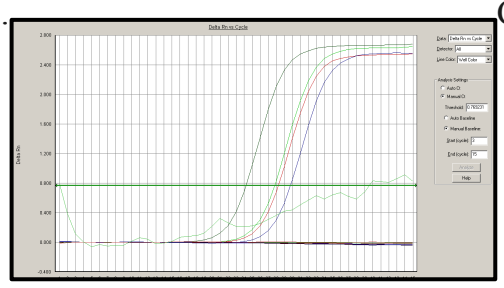
map unannotated retroviral elements in large genomic datasets; we discovered one endogenous, single and type C, LTR in position ch14 {67,691,944 - 67,693,421} just 542bp upstream from our insertion site (Brief description ATTAAA; U5NN:0.37; GT: 0.36;U3NN:0.91;TATAA:0.84; MEME50:0.6; Mot1: 0.39; Mot2: 0.39; Trans:0.36; CpG:0.4; Spl8 :0.62; ttttttttc/tg<>ca/tttttttc). It is not unlikely that our PCR amplified the particular U3 region present in genomic DNA; however, I was not able to map the CF1-related U3 primer to this reference sequence.

Figure 6.2.3 The assessment of mouse DNA contamination in the integration PCRs

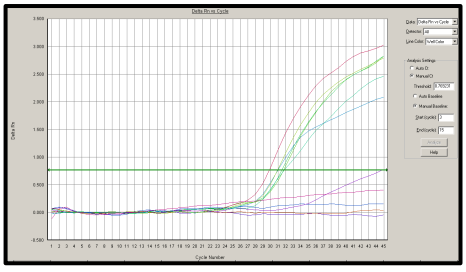
A .

						QC
TFRC	A1	CF1 35 ng	VICTAMRA	Unknown	28.2768	OK
	A2	YAC-I 50ng	VICTAMRA	Unknown	27.8122	OK
	A3	huDNA 10ng	VICTAMRA	Unknown	Undetermined	OK
	A4	293T 35ng	VICTAMRA	Unknown	Undetermined	OK
	A5	293T 50ng	VICTAMRA	Unknown	Undetermined	OK
	A6	j2 15ng	VICTAMRA	Unknown	29.7838	OK
	A7	293T Cf1 50ng	VICTAMRA	Unknown	Undetermined	OK
	A8	J2 Cf1 35 ng	VICTAMRA	Unknown	24.2474	OK
	A9	ARPE 35ng	VICTAMRA	Unknown	Undetermined	OK
	A10	ARPE Cf1 35ng	VICTAMRA	Unknown	Undetermined	OK
	A11	H20	VICTAMRA	NTC	39.5688	OK
PDH	C1	CF1 35 ng	VICTAMRA	Unknown	Undetermined	OK
	C2	YAC-I 50ng	VICTAMRA	Unknown	Undetermined	OK
	C3	huDNA 10ng	VICTAMRA	Unknown	31.747	OK
	C4	293T 35ng	VICTAMRA	Unknown	31.5236	OK
	C5	293T 50ng	VICTAMRA	Unknown	29.877	OK
	C6	j2 15ng	VICTAMRA	Unknown	Undetermined	OK
	C7	293T Cf1 50ng	VICTAMRA	Unknown	31.0034	OK
	C8	J2 Cf1 35 ng	VICTAMRA	Unknown	44.8107	OK
	C9	ARPE 35ng	VICTAMRA	Unknown	30.9615	OK
	C10	ARPE Cf1 35ng	VICTAMRA	Unknown	31.8501	OK
	C11	H20	VICTAMRA	NTC	Undetermined	OK

B .



C.



D.

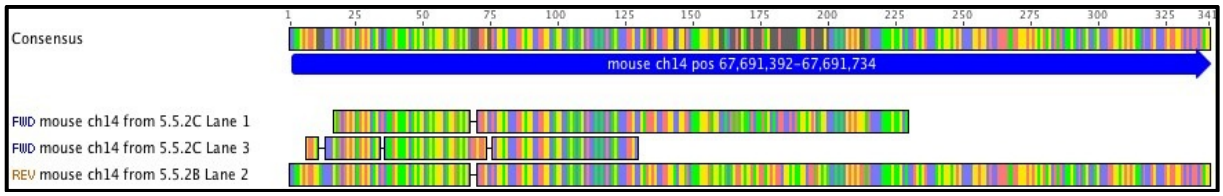


Figure 6.2.3: The assessment of mouse DNA contamination in the integration PCRs

A. Real time PCR results for varying concentrations of genomic DNAs used in splinkerette PCRs assayed for mouse reference gene (Transferrin Receptor TFRC) and human reference gene (PDH) with a dually labelled fluorescent Taq-Probe (AB7500, ABiosystems, UK)**.** QC: quality control**.**

B. Cycle threshold amplification curves for the mouse TFRC reference gene products.

C. Cycle threshold amplification curves for the human PDH reference gene products including a human DNA positive control. The upper cut-off value is CT value 39**.**

D. A complete alignment from genomic sequences identified as possible CF1-MLV integration sites in two different assays and three different samples: in the alignment of the derived sequences from direct Sanger sequencing first is the uninfected J2 gDNA product from Gel A (Figure 6.2.2) Lane 1, second is the uninfected 293T cell gDNA from Gel A (Figure 6.2.2) Lane 3 and third is the infected NIH3T3 gDNA (CF1-MLV) product in Gel B (Figure 6.2.2) Lane 2. The integrations were mapped to mouse chromosome 14, Acc No NT_039606, cytoband D1; position 67,691,392-onwards and the alignment demonstrates their complete identity.

6.3 Next generation sequencing of XMRV integration sites in 22rv1 cells

6.3.1 Construction of library and analysis of the results

Following the failure of conventional sequencing to produce efficient results for integration events in mouse and human cells as per the previous section, we further constructed a DNA amplicon library using the Nextera XT DNA preparation kit for submission to next generation sequencing, using paired end reads of 150bp sequenced in an Illumina (MiSeq) platform (Hammersmith MRC Genomic Unit, Imperial College). For this experiment a positive integration control arising from the splinkerette integration PCR of XMRV producing 22rv1 cells was used. The library was assessed for quality using BioAnalyser at the preparation stage prior to ss DNA normalisation. The amplification of the positive control 22rv1 cell genomic DNA for XMRV integration sites produced a library of 2,426,616 paired end reads. The algorithm used to access these integration sites is described in Figure 6.3.1.

Figure 6.3.1 The analysis algorithm for the confirmation of NGS sequences

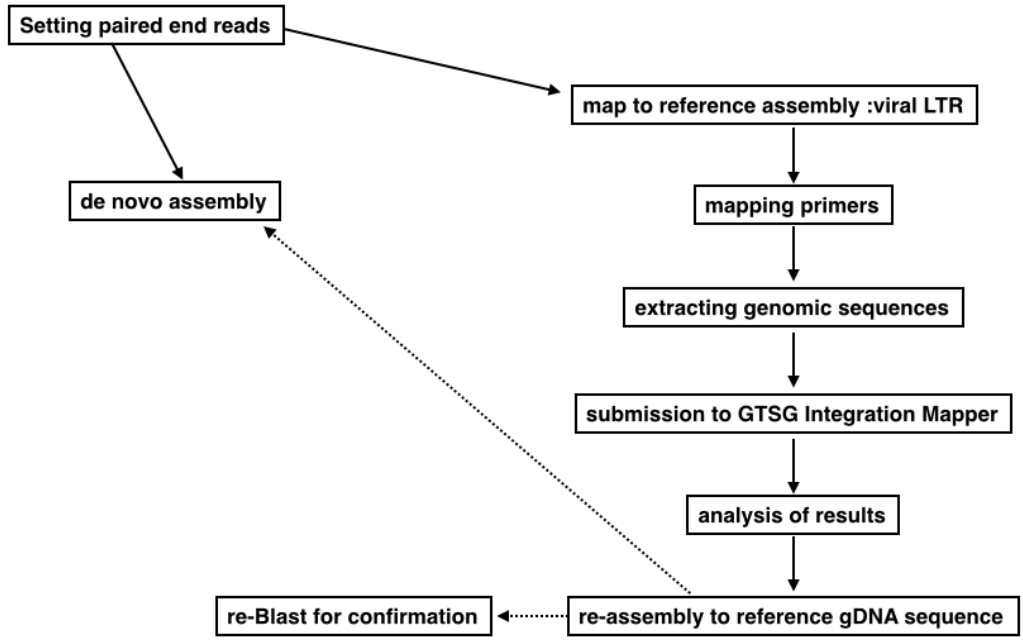


Figure 6.3.1 The analysis algorithm for confirmation of integration sequences in NGS

The schematic briefly describes how the analysis of the reads obtained from the DNA pools submitted to the MiSeq platform for NGS, were subsequently analysed. Functions were performed using the software Geneious, version R8 (BioMatters, USA). In detail, paired end reads are matched automatically, mapping to reference U3-LTR is using Bowtie (2.0). The GTSG Integration Mapper program, QuickMapper is a retroviral integration analysis tool provided by the Gene Therapy Safety Group (www.gtsg.org).

Statistical analyses performed by the QuickMapper used to identify an integration sequence as TRUE or FALSE in the Blast search analysis are explained below:

1) **Blast success**: A filter used to confirm if the blast-search delivered at least one blast hit. 2) **Sequence length filter**: a minimum of 11bp is set and the sequence is true, if it is longer than the specified min-value. 3) **E-value threshold**: true, if the integration sequence has an e-value, smaller than the specified threshold (1E+00). 4) **Identity filter**: true, if the alignment of the integration sequence to the hosts genome sequence has a higher identity, than the specified threshold (set to 0). 5) **Score filter**: true, if the blast score is higher than preset threshold score=0.5 (A measure of the quality of the alignment considering base pair by base pair analysis and introduced gaps). 6) **Ambiguity**: An integration sequence was mapped as true, if the score of the secondary blast hit (in percent of the primary score) does not exceed the preset percentage threshold (set to 0.9).

6.3.2 Results

Using the algorithm, 237 unique human DNA integration sequences were discovered. A full list is found in the Appendix 7. However, only 32 had significant coverage in reads (>50 reads) and were deposited in GenBank (Accession No MG525138-69) An example submission is provided in Appendix 8. This is a lower number of integration sites compared to the dataset of 316 published in acutely XMRV-infected DU145 cells in Kim et al, 2010. The full integration site evaluation of the chronically infected 22rv1 cell line, however, apart from the demonstration of >10 integration events in Metzger et al, 2010 via southern blotting, has not been previously reported.

The integration details of XMRV in the 22rv1 cell line are shown in Figure 6.3.2A for the prevalent integration sites (> 50 reads, 5.2% of total integrations, mean depth of coverage 608X range 58-2201). To determine integration site selection bias, we constructed an *in silica* reference set of 1,000,000 random positions in the 24 human chromosomes of the human reference genome (RefSeq ens62) and used chi square statistics (or Fisher's exact test for counts smaller than 10) to determine differences in

proportions. We also downloaded publicly available datasets for integrations in human cells by MLV, HTLV-1 and HIV-1 as well as the DU145 cell XMRV acutely- infected integrations by Kim et al, 2010 for further comparisons (Table 6.3). The results were centred according to previous reports on transcription start site preference, CpG islands and DNAse hypersensitive regions, but information such as vicinity to LTRs, regulatory features and repetitive elements was also collected. As expected, XMRV in 22v1 cells (all 237 integrations) strongly favoured active transcription sites within 2Kbp (Figure 6.3.2B) and the frequency was a lot higher than for MLV integrations or the XMRV integrations in acutely infected DU145 cells ($P < 0.0005$). Previous results have shown that MLV and XMRV integration favours CpG islands (Kim et al, 2010, Wu et al, 2003). Within the 2Kbp window of the CpG island midpoint, XMRV showed preference for the CpG vicinity compared to the *in silica* dataset ($P<0.0001$) (Figure 6.3.2C). Distance to LTRs was also not found significantly correlated as expected (Figure 6.3.2.D).

Table 6.3 Datasets used for the comparison of integration sites

Integration site datasets used in this study			
Data set (number of integration sites)	Cell type	Reference	GenBank Accession
XMRV (316)	DU145	Kim et al 2010	EU981292- EU981799
MLV(1026)	HeLa	Wu et al, 2003	AY51855 - AY516880
HIV-1	H9, HeLa	Wu et al, 2003	AY516881-AY517469
HTLV-1 (541)	HeLa	Dress et al,2007	EF580177-EF580913

We further extracted the 5-end, 5(five) bases of the human genomic sequence to test against the reported motif of the target site duplication at the virus-host DNA junction (Figure 6.3.3A). A weak palindromic sequence is reportedly unique for each class of retroviruses (Wu et al, 2005). For example, the consensus integration sequence for HIV-1 and MLV are 5′-GTWAC and 5′- VTAB, respectively (using standard International Union of Biochemistry base codes: B # =# C, G or T;V#=# A, C or G;W#=# A or T), while Kim et al, 2010 describes a 5-CTVB motif based on the analysis of 472 integrations from the acutely XMRV-infected DU145 cells. A further difference between our dataset and that in Kim et al, 2010 is that the latter includes patient prostate samples that were discredited by Garson and Towers as identical to cell line integrations consistent with contamination (Garson et al, 2011).

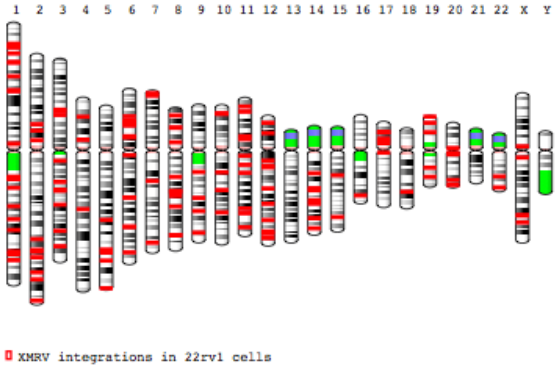
The motif better fits the MLV known consensus of 5'-VTAB and fits statistically (chi square test $p<0.001$) with a 5'-ATAB consensus. Our prediction of 5'-ATAB is based on a 237 sequence dataset with a significant statistical power; however, we are limited by the fact that we are only able to analyse the 5'end of the genomic-LTR junctions and cannot confirm our findings on the 3' end, due to the design of the splinkerette assay.

Figure 6.3.2 Integration sites of XMRV in 22rv1 human prostate cancer cells and their chromosomal locations

A.

chr	position start	or	contig	cytoband
1	23,880,707	-1	AL021154.1	p36.12
1	41,181,728	1	AL031289.1	p34.2
2	85,186,657	-1	AC022210.8	p11.2
2	238,391,628	-1	AC112721.3	q37.3
3	169,490,542	-1	AC078802.14	q26.2
7	73,795,403	1	AC005015.2	q11.23
7	99,068,786	1	AC073063.12	q22.1
7	99,031,933	-1	AC090113.5	q22.1
8	81,489,634	-1	AC009871.21	q21.13
8	128,076,496	-1	AC020688.7	q24.21
9	131,615,624	1	AL441992.17	q34.11
10	63,798,154	-1	AL671972.8	q21.2
11	69,452,209	-1	AP001824.5	q13.3
12	51,626,432	-1	AC139768.14	q13.13
12	119,005,478	-1	AC087880.24	q24.23
14	38,037,962	1	AL121790.4	q21.1
15	85,290,851	1	AC012291.11	q25.3
16	80,712,207	1	AC092332.4	q23.2
16	80,869,037	-1	AC009070.12	q23.2
16	89,741,842	1	AC010538.9	q24.3
17	19,533,650	-1	AC115989.6	p11.2
17	27,170,217	-1	AC024267.18	q11.2
19	39,328,603	1	AC008982.5	q13.2
X	53,341,385	1	AL161779.32	p11.22
X	66,745,723	-1	AL049564.11	q12
X	128,796,217	-1	AL022162.1	q26.1

B.



C.

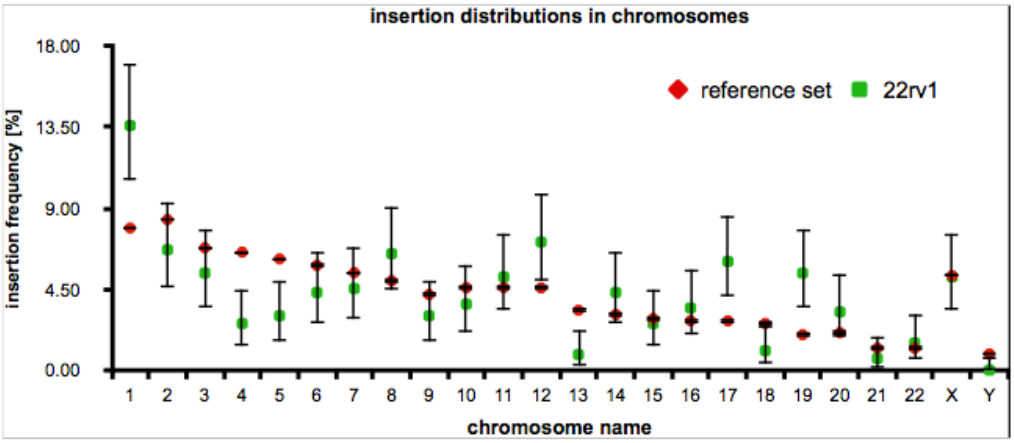


Figure 6.3.2 Integration sites of XMRV in 22rv1 human prostate cancer cells and their chromosomal locations

A. **The most abundantly amplified integration sites of XMRV in the human prostate cancer cell line 22rv1** (average 608X depth of coverage, range 58-2201). Information about the chromosomal location, position start, orientation and cytoband, (Acc Nos MG525138-69) of the corresponding contigs are provided.

B. **Ideogram showing the chromosomal locations** of the 237 discovered integration sites of XMRV in 22rv1 cell gDNA (red bars).

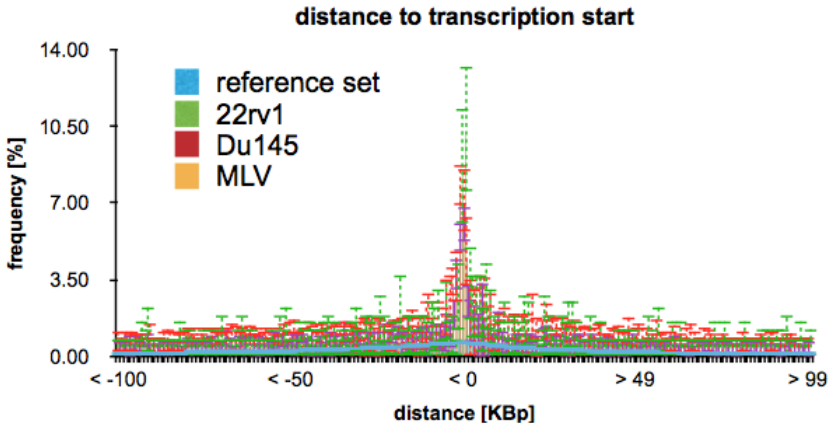
C. **The frequency of integrations per chromosome** is mapped against a random reference set (1million in silica random chromosomal sequences) for statistical evaluation.

Figure 6.3.3 Characteristics of XMRV integration sites in 22rv1 cells

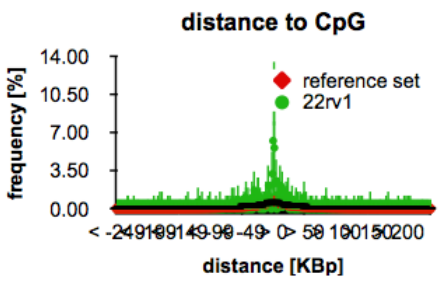
A

		A	T	A	B
	-4	-3	-2	-1	0
A	25	29.6511627906977	25.1937984496124	38.1782945736434	21.8992248062
C	21.3178294573	24.6124031007752	20.7364341085271	17.2480620155039	25.1937984496
G	28.4883720930	24.6124031007752	21.5116279069767	22.093023255814	25.9689922480
T	25.1937984496	21.1240310077519	32.5581395348837	22.4806201550388	26.9379844961

B.



C.



D.

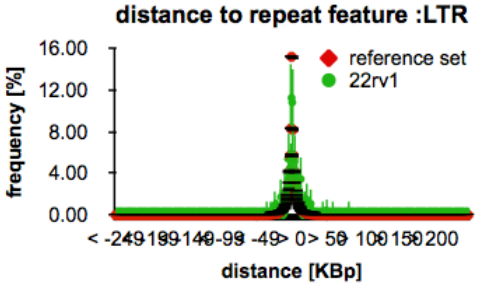


Figure 6.3.3 Characteristics of XMRV integration sites in 22rv1 cells

A. The weak palindromic sequence at the 5'end of the 22rv1 cell line XMRV integrations Distance from integration site is indicated by the numbers -4, -3,-2,-1 and 0, while cells contain the relative frequencies in % of bases discovered in the 283 integration sites in this experiment. The result fits a 5'-ATAB model. This is not in accordance with the described CTVB model for XMRV by Silverman (Kim et al, 2010). Although our dataset is smaller (283 vs 472 integration sites) it does not contain duplicates (14 prostate cancer patient samples in Kim et al that were finally matched with cell line derived samples) while our cell line is a chronically and not acutely infected cell line. Green cells indicate chi square test $p<0.001$.

B. The distance to a transcription start site (TSS) for XMRV integration sites in 22rv1 cells (green) is depicted in comparison to XMRV integration sites in acutely infected DU145 cells (red) and MLV integrations in HeLa cells (orange). The datasets used are described in E. The results are statistically evaluated with a Pearson's correlation test and compared to a random reference set of *in silica* contracted chromosomal locations (blue). The preference for the location in the start of TSS sites is statistically significant for all three viral integration datasets tested.

C. The distance to CpG islands for XMRV integrations sites in 22rv1 cells (green) is graphed in comparison to a reference dataset (red). The integrations are statistically significant for distance to CpG islands compared to the reference control.

D. The distance to repeat feature LTR for XMRV integration sites in 22rv1 cells (green) is graphed in comparison to a reference dataset (red). The integrations are not statistically significant for distance to LTRs compared to the reference control.

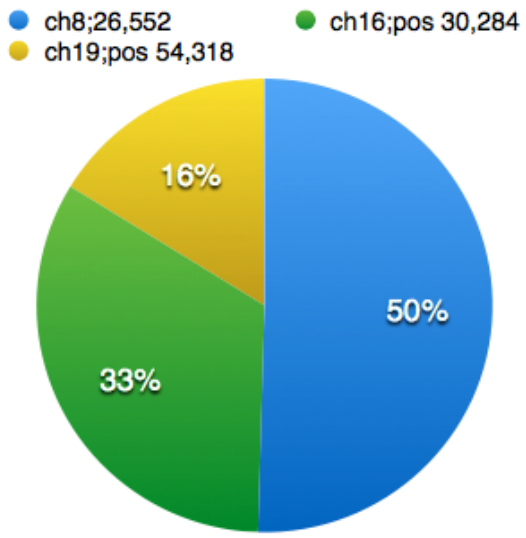
6.4 CF1 cell-related MLV integration in CF1 cells by NGS

Using the same algorithm as described above (Figure 6.3.1) , the amplification of the CF1 cell genomic DNA splinkerette PCR for CF1-MLV integration sequences produced a library of only 2,000 paired end reads containing possible CF1-MLV integration sites. We identified 3 main integrations in ch8-ch16-ch19 for CF1 cell line gDNA (Figure 6.4). An initial striking observation was that the results obtained in this experiment are not in accordance with the direct sequencing result we got from the same gDNA at chromosome 8 pos. 148,187-148,317.

It was equally evident that the amount of reads obtained for CF1 genomic DNA were suboptimal in number (2,000 versus at least ten times more expected) and thus this investigation was revisited on a second occasion in Chapter 7, when we submitted a larger number of CF1 cell related co-cultivations grown for up to 52 days in serial passage to integration next generation sequencing.

Figure 6.4: Mouse chromosomal integrations for CF1-MLV in CF1 cells via NGS.

A.



B.

local Blast conf	sequence	chr	position	reads via LTR map	remapped chr	reads re-assembled	LTR?	adaptor ?	seq with junction	coverage	length	notes
TRUE	ATACAAGAGAC	8	26,552,843	2	conf.	56	8	8	8	30x	130	rfb
TRUE	GGCTATTAACC	16	30,284,445	4	conf.	37	4	0	4	29X	121	rfb
TRUE	ACTTCTTTGATT	19	54,318,638	14	conf.	18	10	2	10	10X	239	rfb

Figure 6.4: Mouse chromosomal integrations for CF1-related MLV in CF1 cells.

A. Percentage pie of mouse chromosomal integrations for CF1-related MLV in CF1 cells.

B. The details of the CF1-related MLV integrations in CF1 cells via NGS. Information about the chromosomal location, position start, and read numbers obtained as well as characteristics such as presence of the adaptor primer, the virus-genomic junction and coverage are provided.

6.5 Discussion

In this chapter, an analysis of the integration of the MLV derived from CF1 cells was attempted and was validated via a concurrent positive control, i.e. the XMRV producing chronically infected human prostatic cancer cell line, 22rv1. Viral integration analysis via NGS has been previously demonstrated for example for Hepatitis B (Sung et al, 2012). Direct sequencing was demonstrative of the ability to map single integration sites when these were abundantly present, however a significant PCR bias by over-amplification would seem to be the driving force in direct sequencing results compared to NGS. The two sequencing experiments provided results for 22rv1 cells, where NGS managed to amplify 237 different integration sites, including the integration site in Chromosome 2 position 85,040,544 that had been uniquely identified from direct Sanger sequencing. For CF1 cells the sequence obtained through direct sequencing was that of a xenotropic/polytropic MLV in a location not seen through later NGS nor detected via x/p RT-PCR. It is true however that our experiment failed to produce a high number of reads to validate our characterisation.

We also managed to effectively develop our integration PCR method as well as experimentally test various gDNAs for integration events. As was expected, a significant contamination presence of mouse gDNA was discovered. As we had controlled for mouse contamination in our molecular method by using mouse-DNA free extraction columns, the existence of sequences in other reagents proves evident. To our knowledge several polymerases also contain mouse gDNA (Tuke et al, 2011) which are more than likely the source of the contamination, and this issue was not further pursued.

We also analysed for the first time the sequences of the integration of XMRV in 22rv1 cells and deposited them in GenBank (Acc Nos. MG525138-69). Although others have mapped XMRV integration events via next generation sequencing, this is the first work that has done so in 22rv1 cells that are the originally infected cancer cell line. By using widely accepted standards of analysis, such as an artificially randomly created,

dataset of 1 million integration sequences and HIV-1, MLV and HTLV-1 integration datasets, we checked the integration sites of XMRV in 22rv1 to discover that the virus exhibits the expected preference for transcriptional start sites and close proximity to CpG islands as hotspots for its integration. This analysis served mainly as a positive control for the discovery of the CF1-MLV integration in CF1 cell gDNA, which proved more troublesome as the CF1 cell line gDNA amplicon from the splinkerette did not amplify effectively, giving us a minimum amount of reads to support a valid finding.

Chapter 7
CF1-MEF and human corneal stem cell co-cultivation

7.1 Introduction

The recent emergence of a recombinant murine retrovirus (XMRV) as a result of human and mouse cell co-cultivation; a completely novel virus that has acquired the capacity to enter and acutely and chronically infect human cells; and its relative presence in a variety of mouse fibroblasts and their respective mouse strains; has now reinstated a more serious understanding over the importance of xenograft and xenotransplantation risk assessments.

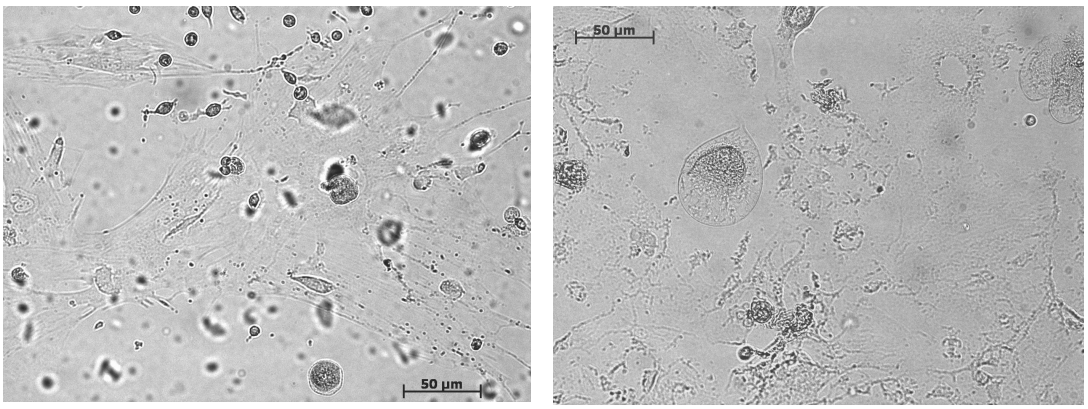
In view of the above considerations, we decided to replicate the clinical protocol used in the expansion of limbal stem cells in the UK (Shortt et al, 2005, Pellegrini et al, 1991) using our virus expressing MLV cell line CF1 and assess the potential of infection of the human limbal stem cells with a wide array of tools, thus informing our assessment over the admittedly rare possibility of a window of transmission of MLV viruses onto human patients.

7.2 *In vitro* expansion protocol for limbal stem cells

Human corneal epithelial stem cells (HCEC) are >80% limbal progenitors and were cultured as described in Materials and Methods (Section 2.2) for 4 passages of 14 days each, adding up to a culture period of 52 days. Limbal stem cells were assessed via microscopy and immunofluorescence every 2 days and at the end of each 14 day passage respectively. Stem cells were assessed for colony formation, colony numbers and differentiation characteristics (smoothness of colony formation, aberrant growth, integrity and confluence of feeder layers) as shown in Figure 7.1.

Figure 7.1 Human limbal stem cells expanded on CF1-MEF cell layers

A. B.



C. D.

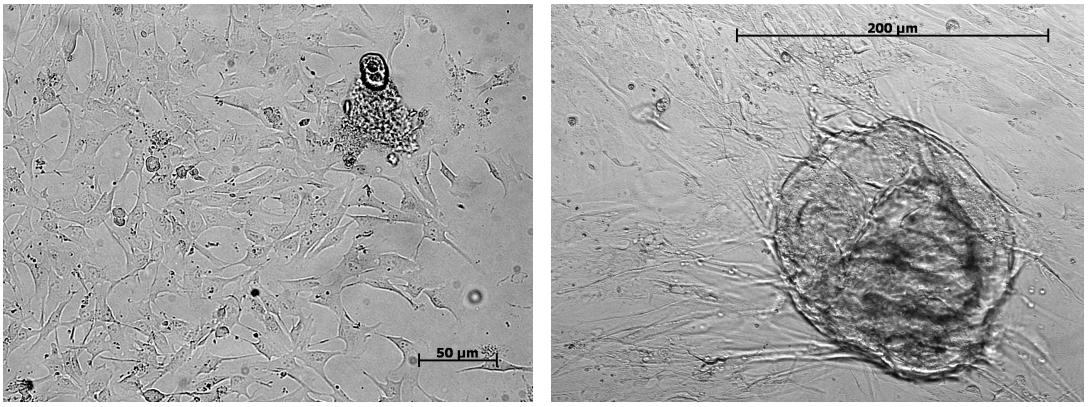


Figure 7.1 Human limbal stem cells expanded on CF1-MEF cell layers

A. Stem cells at day 1 post seeding are seen as large round cells in suspension on top of the mouse feeder layer (CF1-MEF) that are already adherent to the bottom of the well.

B. On day 5 post seeding the stem cells have settled on top of the mouse feeder cells (CF1-MEF) and have started to aggregate, forming 2-3 cell colonies of 50um diameter.

C. On day 10 the stem cells have expanded in number and are numerous on top of the mouse feeder layer (CF1-MEF) while they are found mainly in larger aggregates (>20 cells) during the formation of a stem cell colony (>50um).

D. The appearance of larger stem cell colonies expected to contain >500 cells is shown in day 14 post-seeding.

7.3 Detection of MLV in the co-cultivation of human stem and mouse feeder layers

Supernatant from the co-cultivation of human and mouse cells was harvested over a period of 4 serial passages of 14 days each. The details of the co-cultivation protocol can be found in Methods (Section 2.2) and growth medium (SHEM) recipe in Appendix 5. The PERT assay was performed as described in Methods 2.5.7, and the results for passage 1 and 2 are shown in Figure 7.3.1. The results are indicative of a low RT activity throughout the 4 passages. In addition, RT activity for passage 3 and 4 was less than $1x10^3$ pico units/µl (data not shown).

In MLV *gag* RT-PCR however, there appears to be a significant increase in the level of *gag* RNA expression in the co-cultivated cell supernatants (Figure 7.3.2). This is significantly lower in negative controls such as human stem cells alone and CF1 cells; and appear to be higher for later passages post passage 2, day 28. However, this is supernatant from co-cultivates and as such expected to be high in mouse cell content. We sought to answer this question of mouse contribution to the MLV *gag* expression in the following section 7.5.

Figure 7.3.1 PERT assay results on human stem and mouse feeder cell co-cultivation

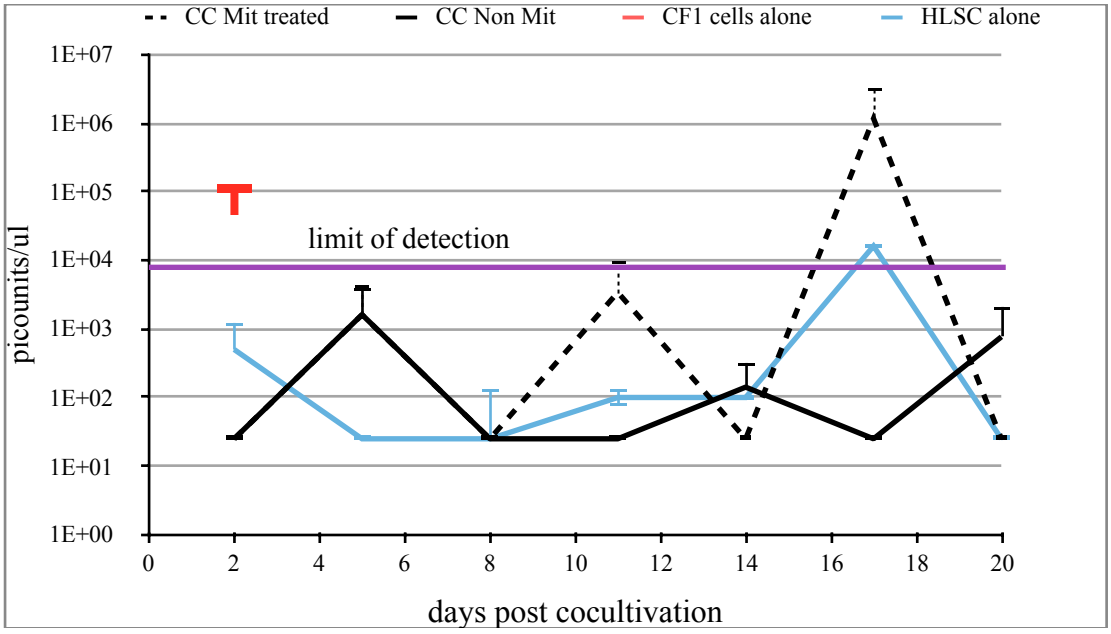


Figure 7.3.1 PERT assay results on human stem and mouse feeder cell co-cultivation

The graph shows the variation in RT activity measurement under and over the limit of detection of the PERT assay (purple line). Only one peak of activity is described at $1x10^6$ picounits/µl for Mitomycin C treated CF1 feeders and human stem cells on the 17th day post co-cultivation or day 3 of passage 2. RT activity for passages 3 and 4 was less than $1x10^3$ pico units/µl (data not shown)

Figure 7.3.2 MLV *gag* PCR assay results on human stem and mouse feeder cell co-cultivations

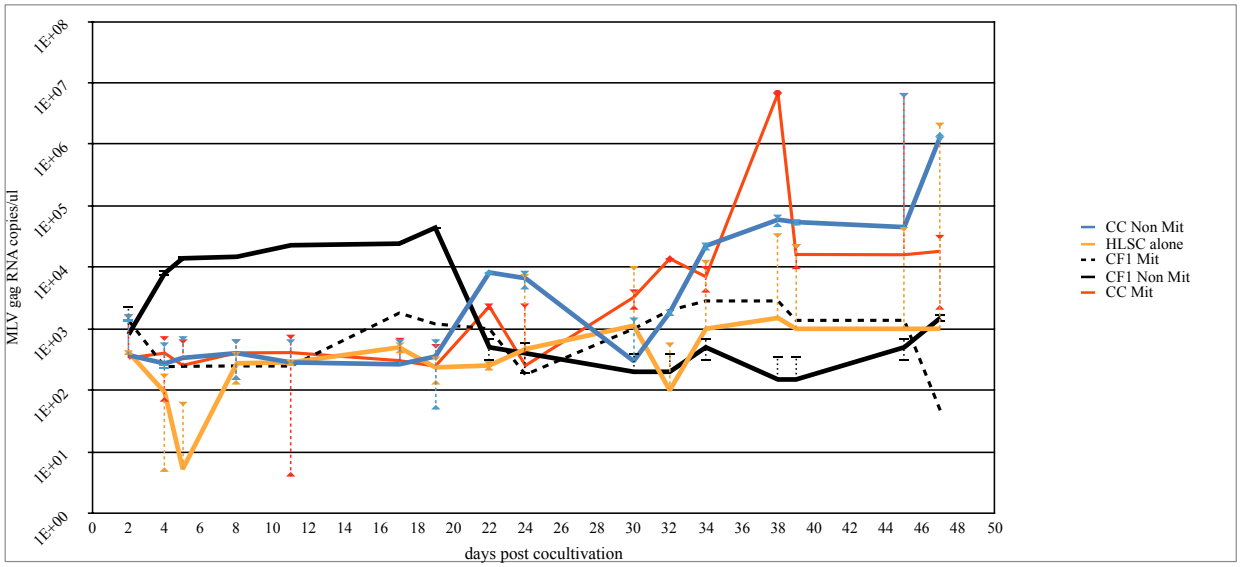


Figure 7.3.2 MLV *gag* PCR assay results on human stem and mouse feeder cell co-cultivations

In red are human stem cells derived (co-cultivated with mouse feeder cells treated with Mitomycin C) total nucleic acids in MLV *gag* PCR (that controls for RNA content via the inclusion of an RT-dependent step). In blue is the same human stem cells' NA via co-cultivation on non-Mitomycin C treated feeders. The negative control (orange) consists of human stem cells growing separately alone on keratinocyte SFX medium. For reference, the mouse feeders alone (black and dashed black) don't exceed $1x10^5$ RNA copies/µl. Error bars are +/- SD for three cell culture replicates per reaction.

7.4 Immunofluorescence staining on co-cultures with anti-MLV Gag

Immunofluorescence staining of anti-MLV Gag was performed on co-cultivated human stem cells and CF1 mouse feeder cells according to a standard protocol described in detail in Methods (Section , 2.6.1). The cells were additionally stained with a human stem cell specific anti-human Integrin A6 (red) and the results for various time points in co-cultivation are seen in Figure 7.4.1. CF1 mouse cells that are expressing anti-MLV Gag are shown to exhibit a characteristic green punctate fluorescence at positions that are cytoplasmic, which is an expected characteristic of MLV-Gag, where as Anti-Integrin A6 (red) stained human stem cells do not get stained with anti-MLV Gag at cytoplasmic regions. Some cross -reactivity occurs but the levels of fluorescence vary significantly. Antibody-deficient negative controls remained negative for any IF signal.

To control for replication competent MLV infection, we infected human stem cells with XMRV from 22v1 supernatant under the same conditions as described above (Figure 7.4.1) and the results of this immunostaining are shown in Figure 7.4.2.

Figure 7.4.1 Immunofluorescence staining of anti- MLV GAG in human stem cells grown on CF1 feeder layers

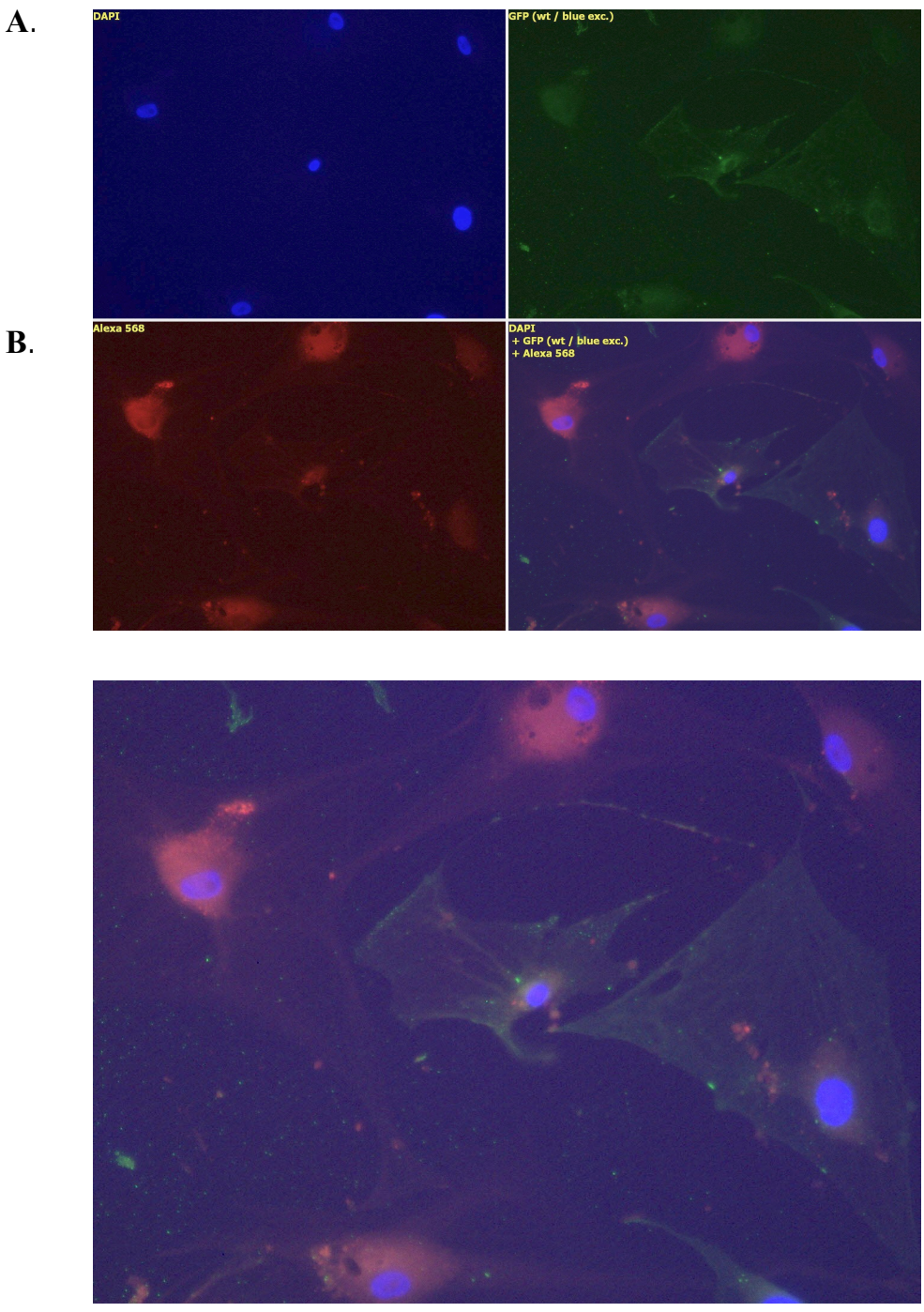


Figure 7.4.1 Immunofluorescence staining for MLV GAG in human stem cells grown on CF1 feeder layers

The cells were stained at 9 days post-first co-cultivation and the results are shown as:

A. Human stem cells and mouse feeder cells (CF1-MEF) stained with DAPI nucleolar stain (blue).
B. Human stem cells and mouse feeder cells filtered at the red spectrum at 648 nm for the Anti-Integrin A6 human protein.
C. Human stem cells and mouse feeder cells filtered at the green spectrum at 488 nm for the anti MLV- Gag viral protein.
D. Human stem cells and mouse feeder cells for both red and green, demonstrate localisation of anti MLV-Gag only for mouse cells and lack of co-localisation of the two stains in the cytoplasm.
E. Magnification of the previous image (D). Some interference of the two stains may exist as all nuclei have stained red, however the intensity of the staining appears to be differentiating between heavily stained (human) and less so (mouse).

Figure 7.4.2 Immunofluorescence staining of anti- MLV Gag in human stem cells infected with XMRV

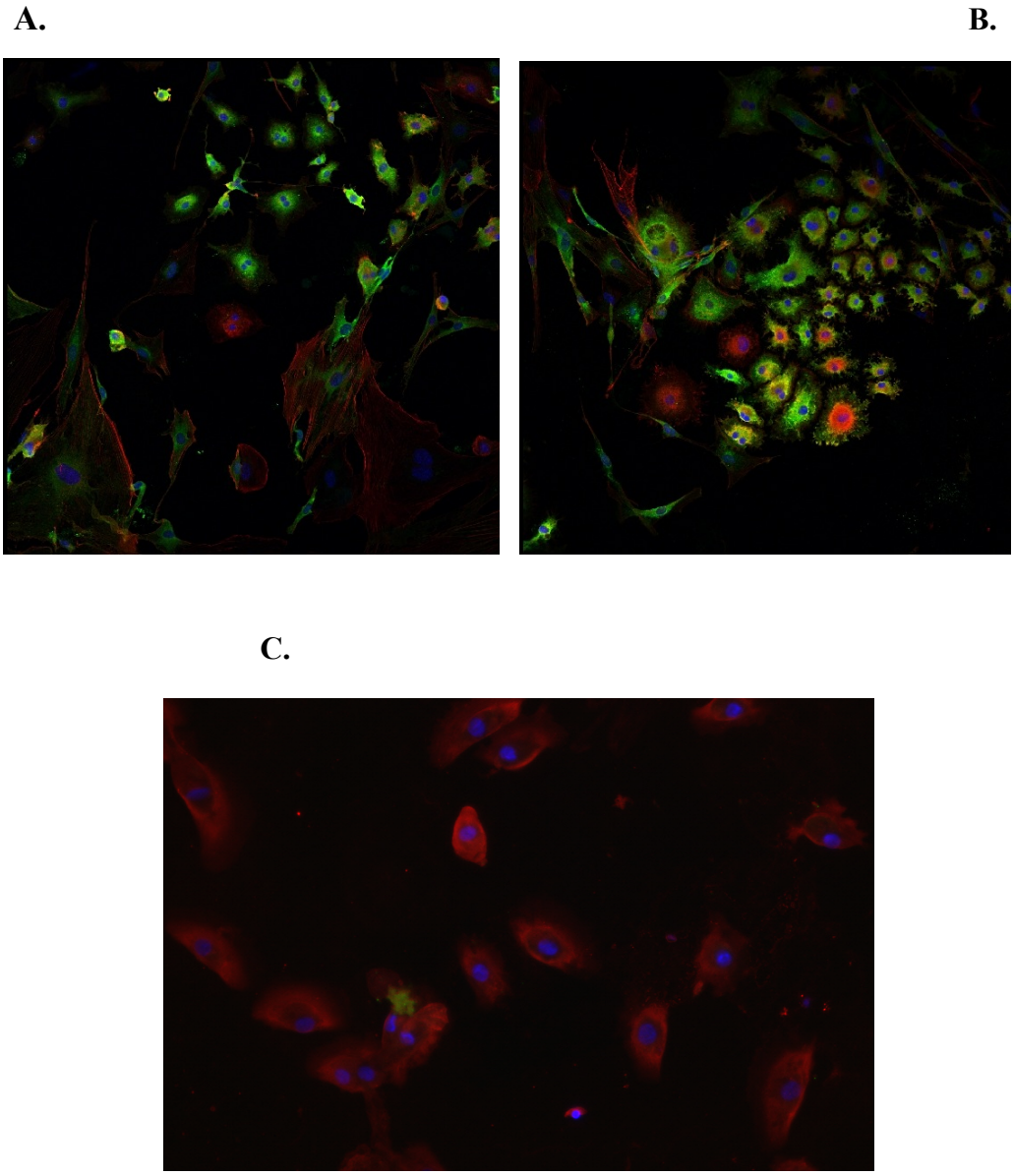


Figure 7.4.2 Immunofluorescence staining of anti-MLV Gag in human stem cells infected with XMRV

A and B show a layer of human limbal stem cells after infection with XMRV. Stem cells were immunostained 7 days post-infection with Anti- Integrin-A6 (red nuclear) anti- K3/K12 which is a differentiation marker for human keratinocytes (red cytoplasmic) and anti MLV-Gag for viral protein (green). The stem cells are visibly infected with XMRV, staining intensely with Anti-MLV Gag (Green) and they are visibly differentiated at approximately 20% (A) and 75% (B) per visual field, having lost cell growth control.

C: Human stem cell uninfected control.

7.5 CF1-MLV *gag* and *env* mRNA expression in total cell RNA from co-cultivations

We extracted total cell RNA from stem cells co-cultivated with CF1 feeder cells and evaluated the expression of MLV *gag* (unspliced) and MLV *env* (spliced) against two reference expression controls, the housekeeping GADPH human and 18SRNA universal RNA expression control. The amplification passed quality control for the positive CF1 and 22rv1 total cell RNA controls, as well as the negative control which consisted of human stem cell RNA alone. We also constructed an artificially mixed total cell RNA serial dilution. To do this we mixed prior extraction, an increasing amount of mouse target cells with a constant number of human cells, to control for the reference and target gene amplification in mixed cell conditions in case of inadvertent assay inhibition.

We tested total RNA extracted from the co-cultivation of human stem cells and mouse feeders for all 4 passages. First we determined mouse content in each sample via ΔΔCT (CT (total RNA)-CT (human RNA)) to estimate the number of remaining mouse cells present in the co-cultivation extracts. In cases where human RNA content was equal to the total RNA, we assumed no mouse contamination. Where there was only total cell RNA but no human mRNA amplification we assumed logically that total cell RNA comprised of mouse cells.

Concurrently we amplified the total RNA extracts with MLV *gag* and *env* primers. The stem cell and mouse cell co-cultivation RNAs produced positive MLV *gag* and *env* RNA signals that were divided by mouse content as determined by the reference RNA controls ΔΔCT described above. The ratio produced was then quantified against the normalisation curve obtained by the artificially mixed RNAs (Figure 7.5). No samples expressed MLV *gag* more than their corresponding mouse content (ratio of 1 and below) while the same was observed for spliced messenger *env.*

Figure 7.5 MLV *gag* and *env* relative quantification in co-cultivated stem cells

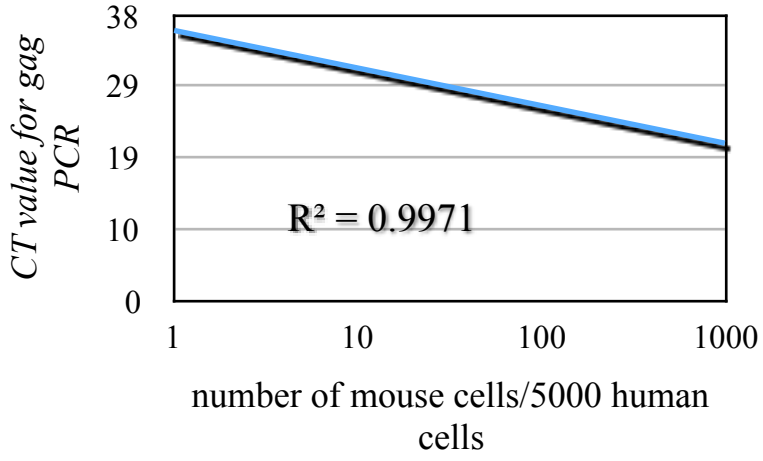


Figure 7.5 MLV *gag* and *env* relative quantification in co-cultivated stem cells

The quantification standard curve obtained for MLV *gag* via serial dilutions of mouse cells in a constant background of human cells. R^2 is 99.7%. This standard curve allowed for the estimation of mouse content in stem and mouse cell co-cultivation extracts (total cell RNA).

7.6 Investigation of CF1-MLV integration in human stem cells grown on CF1 feeder layers via splinkerette and next generation sequencing

7.6.1 NGS strategy and first iteration mapping to reference

We subjected the DNA extracts of the human stem cell and mouse feeder co-cultivations to a protocol for DNA next generation sequencing (NGS) (Methods 2.5.10). We performed the integration "splinkerette" PCR on gDNA extracts from passages 3 and 4 in co-cultivates that were using Mitomycin-treated and Mitomycin-untreated CF1 MEF cells, previously shown to contain e-MLV. The result of the splinkerette is shown in Figure 7.6.1. The purified amplicons from the above reaction were used to construct library pools for DNA-Seq (MiSeq). The obtained reads were analysed according to the strategy followed for RNA-Seq (Chapter 5) with the main difference being that the investigation now focused around a chimeric type of genomic sequence ie. a matched human gDNA and an MLV-viral LTR in linear continuum. This would indicate that the MLV used to infect the HSCs was able to integrate onto human DNA.

The primers used for the splinkerette assay had been initially designed targeting the published RFB virus sequence. An alignment post the full length characterisation of the CF1 related emv-30-like virus (Chapter 6) shows one mismatch at the U3 region between the RFB primer and the viral U3 of emv-30 and xMLV or the osteosarcoma viruses. However, the primers bear the mismatch at the 5' end which is not as detrimental as the 3' end required for primer annealing and so we did not design new primers (Appendix table 1 and 8).

The strategy briefly comprised of a a) reference assembly of paired reads to U3 CF1-MLV b) mapping and annotation of splink2 (adaptor) and CF2 primers c) blasting of sequences found between the splink2 and U3 primers d) reiteration via *de novo* assembly and map to reference assembly against the discovered sequences e) submission to gtsg.org for mapping into chromosomal regions as integration sequences.

The last step (QuickMap) characterises genomic sequences as true integration events according to a series of steps, which are:

Blast success: A filter used to confirm if the blast-search delivered at least one blast hit.

Sequence length filter: a minimum of 11bp is set and the sequence is true, if it is longer than the specified min-value.

E-value threshold: true, if the integration sequence has an e-value, smaller than the specified threshold (1E+00) (extremely stringent).

Identity filter: true, if the alignment of the integration sequence to the hosts genome sequence has a higher identity, than the specified threshold (set to 0).

Score filter: true, if the blast score is higher than preset threshold score=0.5 (A measure of the quality of the alignment considering base pair by base pair analysis and introduced gaps) (very loose).

Ambiguity: An integration sequence was mapped as true, if the score of the secondary blast hit (in percent of the primary score) did not exceed the pre-set percentage threshold (secondary/ primary) (set to 0.9).

Statistical qualities of the acquired sequences were tabulated for quality control as the phenomenon of fewer reads containing the true LTR-genomic junction compared to pure DNA and DNA with splink adaptor was observed. This was considered favouring our hypothesis as due likely to a small percentage of the integration sequences in the cells used to derive the DNA target in the splinkerette assay, but the much higher likelihood of a possible result due to an inadvertent PCR artefact effect was considered main.

Further analysis of the co-cultivation samples revealed several interesting finds. The number of sequences obtained as paired end reads and the results of the first iteration reference mapping to the viral LTRs, either the U3-CF1-MLV or U3-XMRV are shown in Table 7.6.1.

Figure 7.6.1 Splinkerette PCR on genomic extracts from co-cultivated human stem and mouse feeder cells

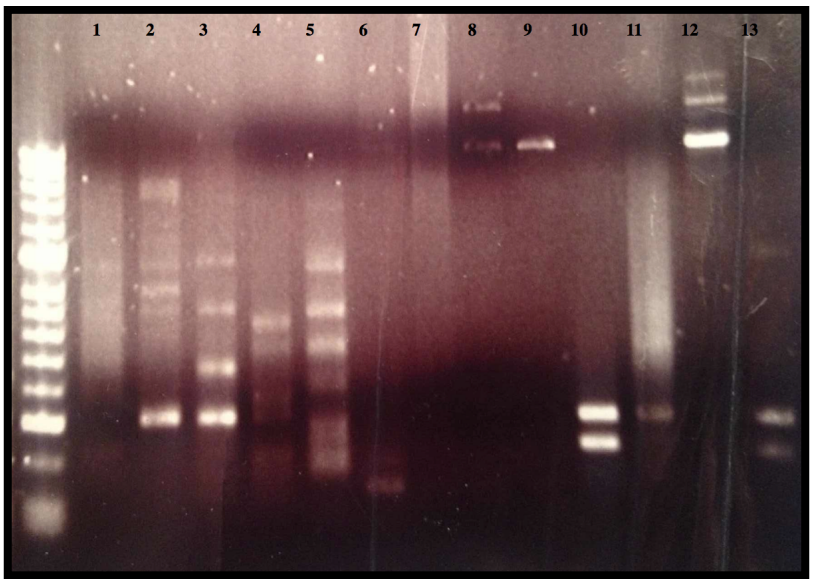


Figure 7.6.1 Splinkerette PCR on genomic extracts from co-cultivated human stem and mouse feeder cells

The splinkerette PCR gel electrophoresis with the amplicons obtained from the amplification of genomic junctions of CF1 MLV and XMRV in several gDNAs and cell culture extracts from stem cell co-cultivations. The lanes in detail show: lane 1-2-3 gDNA from stem cell co-cultivation in passage 3- after Mitomycin treatment of the feeder layer; lane 4-5 gDNA from stem cell co-cultivation in passage 3 without Mitomycin treatment ; lanes 6-7 gDNA from stem cell co-cultivation in passage 4- after Mitomycin treatment of the feeder layer; lanes 8-9-10 gDNA from stem cell co-cultivation in passage 4- without Mitomycin treatment ; lane 11 human stem cell gDNA (negative control) ; lane 12 22rv1 gDNA (positive control) ; lane 13 CF1 genomic DNA (positive control).

Table 7.6.1 Output and mapping to reference reads obtained from integration PCR amplicons

source	sample lane	paired end reads	mapped to viral LTR
CF1 gDNA	13	1372	34
human stem cells	11	797,496	152,463
22RV1 gDNA	12	2,426,616	849,373*
passage 3 Mitomycin treated	1-2-3	1,447,570	165,395
passage 3 untreated	4-5	1,698,864	137,724
passage 4 Mitomycin treated	6-7	291,442	35,338
passage 4 untreated	8-9-10	852,074	80,893
hSCs+ 22rv1 supernatant	ns	2,509,784	32,668*
hSCs +CF1 supernatant	ns	5,389,132	2,398

* XMRV primers were used in the reactions marked with an asterisk / ns=not shown in gel in Figure 7.6.1

Table 7.6.1 Output and mapping to reference reads obtained from integration PCR amplicons

The table briefly shows the number of paired end reads obtained from MiSeq, according to the sample submitted for integration analysis (gel in Figure 7.6.1) as well as the first iteration reads to reference U3 sequences of MLV genomes (RFB or XMRV).

7.6.2 Integration site analysis of co-cultivated stem cells

The output from the results analysis according to the strategy identified above from stem cell co-cultivations and controls on human integrations is shown in Figure 7.6.2.1 for the four main samples from stem cell co-cultivations (passage 3 and 4, with and without Mitomycin). In total, the aligned read-numbers after second step parsing for true presence of the viral LTR-genomic junction, is significantly reduced compared to the initial number of sequences from first mapping, probably due to the low analytical threshold of the "mapping to reference" alignment and gap tolerance. This is the main reason why several iterations were performed as well as a combination of *de novo* and additional re-mapping to genomic DNA reference alignments for confirmation. As the samples submitted were highly variable due to the co-cultivation strategy (where even single cell events would be considered likely), our small number findings underwent further analysis.

The quantitative results of putative integration sites analysis are shown in Figure 7.6.2.1A. After their discovery and subsequent re-iterations and parsing for the presence of the viral LTR-genomic junction, we re-blasted the sequences to identify their origin as per human or mouse genome reference. This was depicted with RED for true human and GREEN for mouse/human sequences ambiguity. We further mapped to reference using viral LTR sequences that would mis-anneal but still amplify with our ecotropic U3-CF1 MLV primers, to look for possible xenotropic integrations present in the samples. Indeed, this strategy was only used to inform of existing sequence variability in the pool of sequences and not used as a potential confirmation of polytropic or xenotropic virus integration (not accounted for by our ecotropic MLV pre-NGS splinkerette primers).

Our results for the first sample from passage 3 of co-cultivated human and mouse cells (with Mitomycin), show one human integration mapping 82 reads to human chromosome 19. After controlling for quality (viral LTR AND adaptor sequence present OR viral LTR-genomic junction present) the depth of coverage for this alignment falls at 13x. From further quality control via the BLAST criteria, the

sequence does not seem to be a human integration site i.e. returns as TRUE/ FALSE for human/mouse reference genome. The ambiguity concerns overlap not with one but many mouse sequences.

Further in the initial screen of passage 3 co-cultivates, were 359 sequences mapping to human chromosome 1, a sequence which also failed the same TRUE/FALSE criteria as above.

From passage 3 of co-cultivates without Mitomycin, one human integration sequence for chromosome 6 reappears with a significant 1,765 reads (not shown in table as explained below). Parsing for viral LTR and genomic sequence continuum without gaps this remains high at 1,765 sequences. However, further careful observation of the alignment demonstrates that the post alignment contig from this particular co-cultivation sample suffers from one variation in nucleotide 14 of the human chromosome 6 (a T to G mismatch) throughout the entire 1,765 sequences. This automatically fails the TRUE human result from BLAST and aligns the sequence 100% with mouse chromosome 8. Indeed, this one mismatch inadvertently shows that these have arisen from the mouse cell content.

Following from the same sample, one further integration is found in chromosome 12 and is TRUE, however the depth of coverage is only two sequences and the viral LTR maps to xenotropic MLV LTR.

From passage 4, with Mitomycin, five (5) integration sequences map as TRUE to human chromosomes (shown in Figure 7.6.2.1A in red) and one (1) overlaps with mouse chromosome (green). However, all of these failed to map with adequate depth at the LTR -genomic DNA junction where minimum copies are observed. The sequences mapped in BLUE are seen using mapping to LTR strategies against xenotropic primers that have several non-detrimental mismatches with our used primers and as such may even more likely have arisen via PCR recombination bias. One of these at human chromosome 12 seems to occur in different origin samples , making it safe to assume that this is likely a contamination or artifactual sequence.

Similarly an integration read (4 reads) in human chromosome 3 at position 141,659,520 appears in subsequent samples fro passage 4 in similar minute quantities at 34 bp apart. Looking at the control sample derived from stem cells infected via CF1 cell supernatant discovers the same human chromosome 3 integration sequence, a further proof that this is a human-specific viral-like sequence or primer misannealing to an LTR-like human promoter sequence and not a novel infection event.

From passage 4, without Mitomycin; the cells again appear to bear multiple integrations, one of which aligns 682 reads with viral LTRs but only 2 with a continuous genomic junction with a TRUE human chromosome X sequence. To demonstrate this altering effect of the lack of viral LTR-genomic junctions in the alignment, an indicative depiction of the above assembly in chromosome X is shown in Figure 7.6.2.1B and the complete alignment of the 2 observed chimeric sequences seen in Figure 7.6.2.2.

The human limbal stem cells infected with XMRV DNA virus from 22rv1 cells, a human virus positive integration cell control demonstrated only one main novel viral integration event in human chromosome 10 (Figure 7.6.2B and C). This was seen in significant amounts of reads demonstrating continuous LTR-genomic DNA junctions (15060 reads) of a mean 600 bp length. To ascertain if this is true novel integration site, we initially used the online program Retrotector (http://retrotector.neuro.uu.se/pub) to look 1000bp up and downstream in the flagged genomic position in the human DNA for endogenous LTRs. There is one single LTR1_S1 of C type genus at position 95289 to position 96135. This, however, does not anneal with the U3 XMRV primers of the splinkerette, suggesting that the infection likely had a hotspot at this integration site, rather than an endogenous element amplification that is also lacking from other uninfected stem cell controls. A clonal integration event of XMRV in human limbal stem cells at human chromosome 10 (pos 112,181,201, contig AL355512.22 cytoband q25.2) mapped via amplicon NGS through our co-cultivation conditions is an original finding.

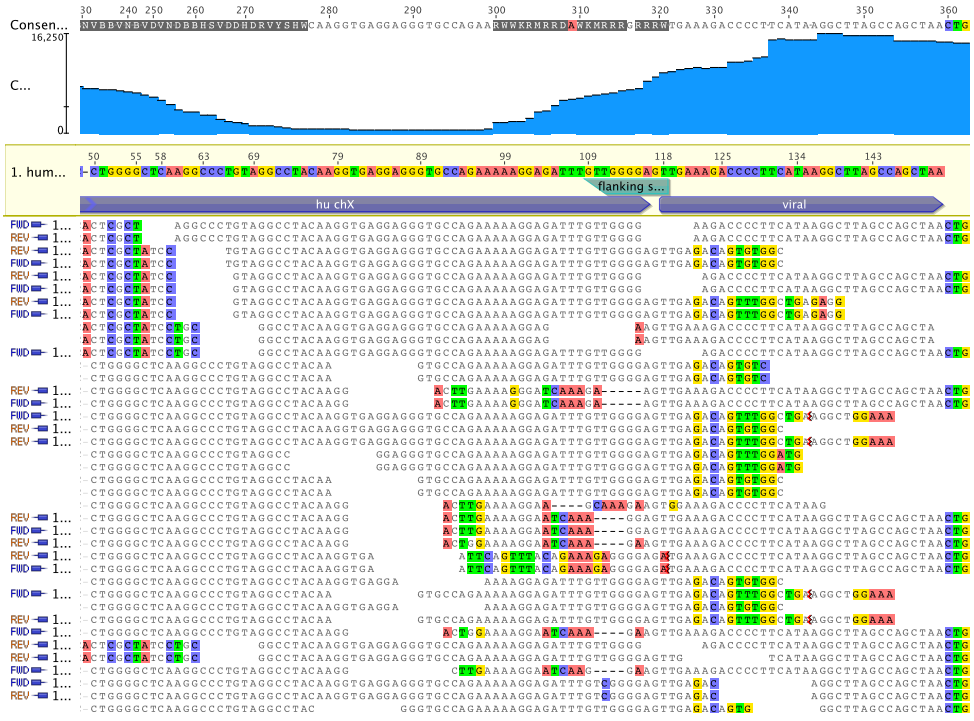
The negative control that comprised of human stem cells alone did not show any integrations that mapped TRUE to human gDNA. Ambiguous reads mapped to human chromosomes (Figure 7.6.2.1A) but indicated cross-coverage with mouse sequences. The integration sequence from mouse CF1 cells (genomic mouse cell control) was also not found in this sample, likely due to the low genomic content of the co-cultivates at the end of the passages with mouse cells. We therefore attributed observed mouse integrations to mouse reagent contamination in some step of the amplification or sequencing, consistent throughout the experiment.

Figure 7.6.2.1 Integration sites from the co-cultivation of stem cells

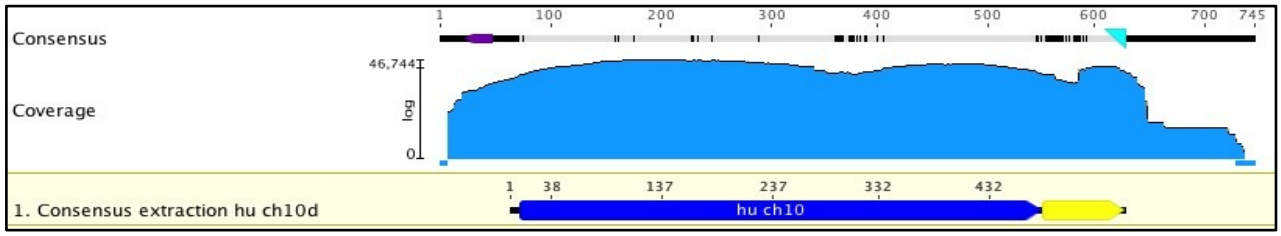
A.

NGS output	chr	position	LTR reads	junction present	chr	position	LTR reads	junction present
Pass 3	Mitomycin				non Mitomycin			
	1	200,284,241	359	69	1	200,284,311	652	296
	19	45,862,952	85	13	12*	51,626,414	2	2
Pass 4	2	3,930,652	325	1	1*	23,880,651	30	14
	1	18,872,218	188	1	8	144,476,891	2	2
	5	142,960,929	489	1	17*	27,170,122	1	1
	16	76,319,494	2	2	X	107,406,158	682	2
	17	59,063,117	150	2	1*	232,841,884	8	3
	1	105,123,383	2	2	3*	137,932,785	1409	2
	12*	51,626,393	6	6	3*	141,659,520	43	9
	3*	141,659,554	4	4	X*	53,341,449	4	4
SCs alone	3	70,100,133	146	13				
	12	65,905,236	197	41			**true human**	
	18	12,459,775	169	57			**true/false 50%**	
SCs with CF1 supernatant	3	141,659,522	180	33			**true human but xMLV mapped alone -primer mismatch**	

B.



C.



D.

mouse ch ambiguity	blast	ch	position	orientation	contig	cb	reads via LTR map	reads re-assembled	LTR?	adaptor	seq with junction	length	mean depth
none	1	10	112,181,201	1	AL355512.22	q25.2	4985	158,021	13081	1108	15060	600bp	23847X

Figure 7.6.2.1 Integration sites from the co-cultivation of stem cells

A. The integration sequence analysis from NGS pool results. In red are sequences that have confirmed human only, in green are sequences that mapped equally to human and mouse DNA and in blue are sequences that were derived after additional mapping to reference xenotropic MLV LTR (that was not used in the original PCR).
B. Mapping passage 4 co-cultivation reads from non Mitomycin treated cells to human chromosome X. The initial genomic sequence is identified in *de novo* or BLAST searches of the NGS pool from the integration experiments- the subsequent alignment from within the sample pool has significant gaps and a true viral integration sequence in direct continuum is only seen in 2/682 of the aligned reads (shown in detail in Figure 7.6.2.2). Most of the observed sequences only align partly with either the viral LTR or the human DNA.
C. The human chromosome 10 integration site of XMRV infecting limbal stem cells showing the high depth of coverage of the alignment. The unique integration sequence in chromosome 10 was likely a clonal event. Fifteen hundred and sixty (15,060) sequences of the 158,021 aligned show true viral integration junctions with human genomic DNA at chromosome 10.
D. Details of the XMRV integration sites analysis in human stem cells.

Figure 7.6.2.2. MLV viral sequences integrated into human genomic DNA

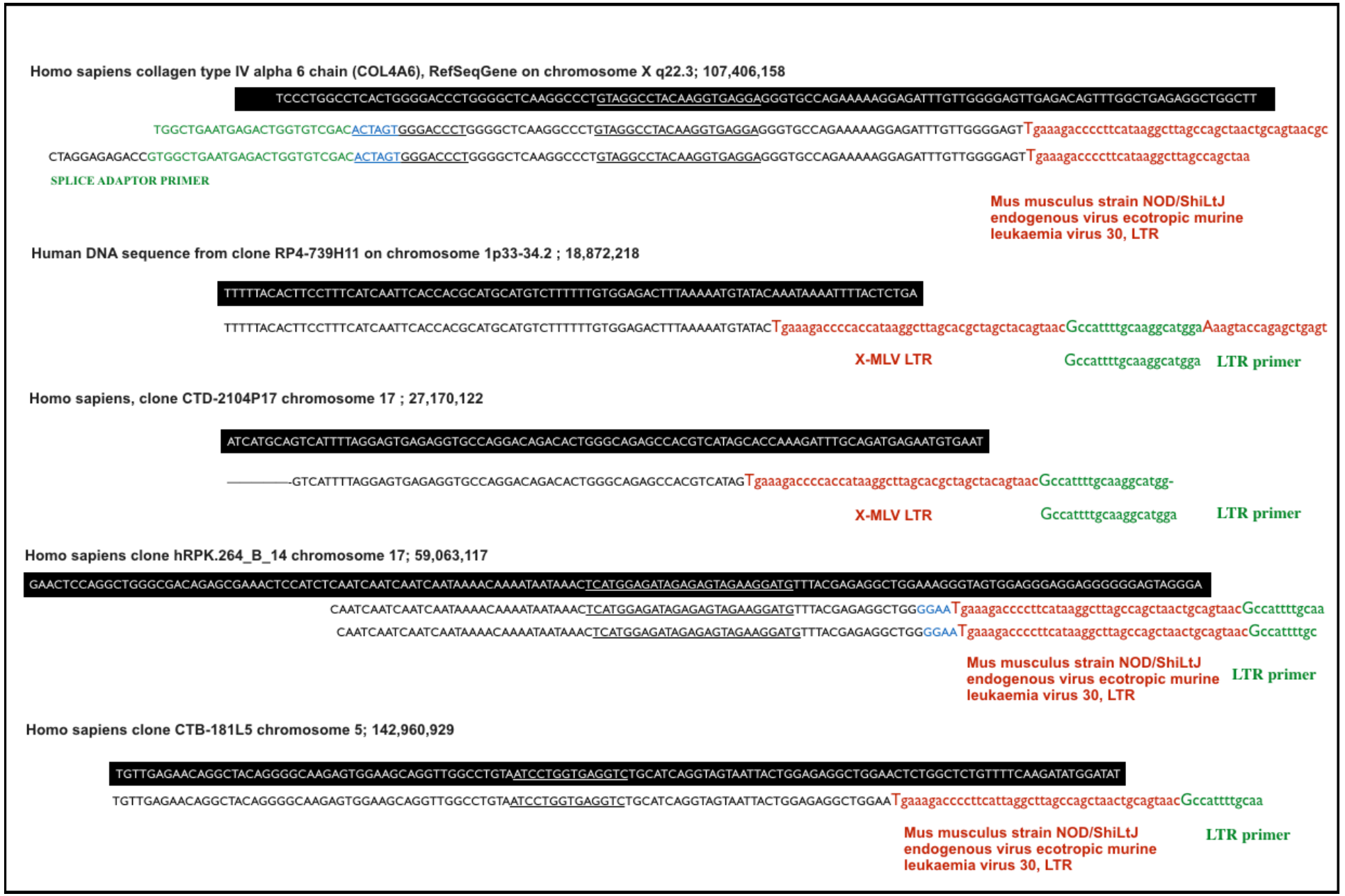


Figure 7.6.2.2. MLV viral sequences integrated into human genomic DNA

The figure depicts the exact alignment of some of the discovered sequences via NGS that support an MLV virus integration into human genomic DNA. The reference genomic sequence is seen in bold black and against it is aligned the hybrid integration sequence with the viral LTR in small red. Any mismatches are seen in blue. In some of the sequences the viral LTR primer used in the splinkerette is seen in green. Any adaptor sequences or primers are seen in either underlined or green.

The integration seen in 2 sequences in Hu chromosome X q22.3; 107,406,158 (top), is also confirmed via PCR with primers (underlined middle) designed post NGS (for details see Section 7.6.3). In Hu 1; 18,872,218 (second from top) only one integration sequence is seen and that maps to a viral LTR of a xenotropic- MLV. Two integration sites were seen in Hu 17; one in position 27,170,122 with a x-MLV LTR and two in position 59,063,117. The last aligns with an ecotropic MLV that fits with the virus from CF1 cells. The last integration site was identically amplified at the confirmatory PCR stage (Section 7.6.3). It is interesting how this is the only site that resembles a true integration site as it demonstrates a flanking sequence of 4 bp between the viral LTR and the human genome. Last, only one sequence is chimeric with an emv-30 LTR and a human chromosome 5 sequence that also confirmed via PCR (following section).

7.6.3 Verification of integration sites defined by NGS

We further examined the validity of our findings by designing primers on the obtained sequences from NGS of integration events, based on the discovered 5'end -genomic DNA from human chromosomes, and testing the original extracts of stem cell co-cultivations by PCR. The results of this amplification are shown in Figure 7.6.3A. The amplicons obtained were further sequenced via direct sequencing to confirm their content. Obtained results are shown in Figure 7.6.3B.

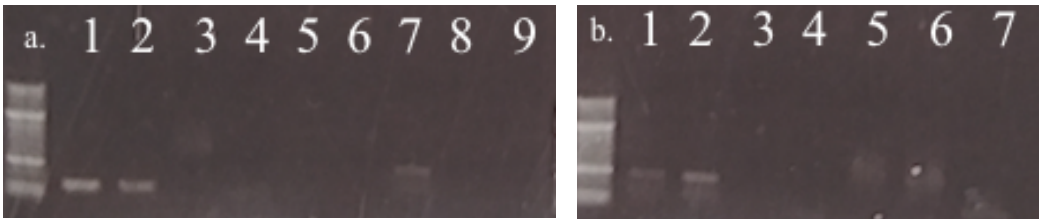
Confirmatory PCR products were obtained for 5 human integration sequences. One was in human chromosome 19 from passage 3 of the human stem cell co-cultivation, with Mitomycin treatment, but this had identified as 50% human and 50% mouse. Three sequences were confirmed on genomic DNA from passage 4 of human stem cells co-cultivated with mouse feeders (Mitomycin treated) on human chromosome 1, human chromosome 17 and human chromosome 5. One sequence was re-identified as valid via direct sequencing from passage 4 without Mitomycin on chromosome X at a position near the *collagen a* gene. All of the above sequences were discovered in minute 1-3 reads as full length sequences containing the 5'-LTR, genomic sequence and their junction with the spinkerette PCR primers in the original pool of DNA reads. This constitutes less than 0.01% of the original number of reads and although may have arisen from PCR artefacts, there is a likelihood that they are true integrations. This is because we confirmed their presence using the original extracted DNA from the stem cell co-cultivations that had not been subjected to PCR. It may be; in the cases where these are true sequences; that the human cells acquired these integrations via single cell infection events or that they have arisen through annealing with promoter LTRs specific to exon regions in an alternate splicing event.

A detailed analysis of the chromosomal positions with distances from active genes, CpG islands and/or other characteristics was performed. None of the viral 5-LTRs confirmed in putative integration sites are of significant length to differentiate integrated MLVs in their origin as ecotropic, polytropic or xenotropic-related. None are on sites of stably integrated pre-existing type-C retroviral elements, but LTRs are

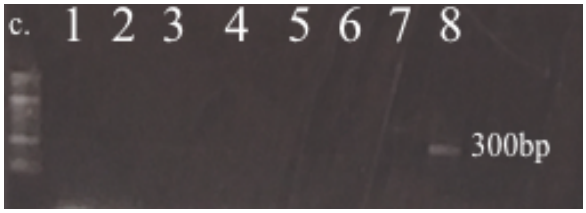
distinguishable MLV. An index of MLV viral LTRs that may mis-anneal effectively, as the original U3 primers are broadly generic can be found in Appendix 9.

Figure 7.6.3 Verification of integration sites from genomic DNA of human stem cell co-cultivations

A.



B.



1. CCCAGTGCAGGGCTGGCCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCAGG
CGGATCACGAGGTCAGGAGATCCAGACCATCCTGGCTAACGCGGTGAAACCCCGTCTCTACTAAAA
ATACAAAAATTAGCTGGACGTGGTGGCACGCACCTATAGTTCCAACTACTCAGGAGCCTGAGACAG
GAGAATTGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCACACCAGCTGCACTCCAG
CCTGGGTGAAAATGAGAGTCTGTCTCAAAAAAAAAAAGAGTTCTTCTGGCAGGGACTAGAGGCA
GGAGACTGGGTGGCAGGCAGGGACCTGAGCAGGAGGGGAGGGGCCTTGGCCAGGATGTGGCCATG
GGGAGAGAGAAAGGTATGGTTTGGAGAGAGATGTGGGAGGTAGAGAGGGAAGGGAGAGAGAAGG
GGGGAGGGTGAAAGACCCCTTCATAAGGCTTAGCCAGCTAACTGCAGTAACGC
2. GTGGCTGAATGAGACTGGTGTCGACACTAGTGGGACCCTGGGGCTCAAGGCCCTGTAGGCCTA
CAAGGTGAGGAGGGTGCCAGAAAAAGGAGATTTGTTGGGGAG(T)TGAAAGACCCCTTCATAA
GGCTTAGCCAGCTAACTGCAGTAACGC
3. GAATGAGACTGGTGTCGACACTAGTGGGATCCCGCCACCTGGACAGTGCAAGACCTCCTGGC
CCCAGGCTGCCTGACTTGCCAGCTTAGGATTCTTTCTTGAGGTAGCCAAGCAAAGTGAGTGC
TGGCTCAGGAGAGGGATGGAGAGAGGCTGGAAAGGGAATCAAAGAAG(T)TGAAAGACCCCTT
CATAAGGCTTAGCCAGCTAACTGCAGTAA
4. ACAGGCTACAGGGGCAAGAGTGGAAGCAGGTTGGCCTGTAATCCTGGTGAGGTCTGCATCAG
GTAGTAATTACTGGAGAGGCTGGAATGAAAGACCCCTTCATTAGGCTTAGCCAGCTAACTGCA
GTAACGCCATTTTGCAA
5. TGTCGACACTAGTGGGATCACTTGAGGTCAGGAGTTCAAGACCAGCCTGGTCAACATGGTGAA
ACCCTGTTTCTACTAAAAATACAAAAATTAGCCTGGTGTGTTGGCGTGCGCCTGTAATTCCAG
CTACTTGGGAGACTGAGGTGGAAGAAACTCTTGAACCCAGGAGGTGGAGGTTGCAGTGAGCT
GAGATCGTGCCACTGAACTCCAGGCTGGGCGACAGAGCGAAACTCCATCTCAATCAATCAAT
CAATCAATAAAACAAAATAATAAACTCATGGAGATAGAGAGTAGAAGGATGTTTACGAGAGGC
TGGGGAATGAAAGACCCCTTCATAAGGCTTAGCCAGCTAACTGCAGTAACGCCATTTTGCAA

Figure 7.6.3 Verification of integration sites from genomic DNA of human stem cell co-cultivations

A. **The gel electrophoresis from the PCR on genomic DNA** obtained directly from human stem cell co-cultivations at various times in passage. In detail,

a. shows the amplicons with primers directed at genomic positions as identified from NGS , as follows: Lane 1. huX (107,406,158). Lane 2. hu19 (45,862,952). Lane 3: hu3 (141,659,554). Lane 4: hu16 (76,319,494). Lane 5: hu2 (3,930,652). Lane 6: hu1 (105,123,383). Lane 7: hu17 (59,063,117). Lane 8: hu1 (232,841,884).

b. as follows Lane 1: hu5 (142,960,929). Lane 2: hu1 (18,872,218) Lane 3: hu3 (137,932,785) Lane 4: hu8 (144,476,891). Lane 5: huX (53,341,449) Lane 6: hu1 (23,880,651). Lane 7: hu17 (27,170,122).

c. as follows Lane 1: hu1 (200,284,241) Lane 2: hu6 (92,411,541) Lane 3 hu1 (105,123,383), Lane 4 hu3*(141,659,520) lane 5 hu1*(232,841,884) lane 6 Hu12* (51,626,414) Lane 7 huX* (53,341,449) and Lane 8 22rv1 positive control. The integration sequences that bear an asterisk were primed with x-MLV specific LTR primers.

B. PCR products from integration confirmation obtained via direct sequencing. Index of colours: in blue is human genomic DNA, in red are integration PCR primers (pre NGS), in green is the confirmatory PCR primer (post NGS), underlined are the adaptor sequences and in black bold the viral LTRs.

1. Hu 19 (ERCC2 gene). We observe a gap of 25bp between the recognised human/mouse genomic sequence (blue) and the start of the viral LTR (black bold) that was not identified in the 85 reads from NGS sequencing. This gap maps as mouse DNA.
2. Hu X at AL031177.1 cytoband q22.3 (Collagen *A* gene). There is no gap between the genomic DNA and viral LTR. The gap after the splink primer is consistent with the adaptor linker sequence. This integration sequence was present in 682 mapped reads, only 2 of which contained the genomic junction.
3. Hu1 at position 18,872,218. There is an 18bp gap between the human gDNA and the start of the viral LTR. This integration site was seen in 1 read.
4. Hu5 at 142,960,929. The splink primer was not found. Seen in 1 read.
5. Hu17 at position 59,063,117. A small gap of 4bp is found between the genomic sequence and the viral LTR. Seen in 2 reads.

7.7 *Mus dunni* infection assay and reverse transcriptase activity for polytropic viruses

The detection and confirmation of human genomic DNA sequences in direct junction with viral LTRs confirms our level of certainty over the NGS experiment's sensitivity and several concerns over the emergence of PCR artifactual sequences.

These minute findings warrant a viral infectivity proof of concept experiment that would indicate that a replication- competent recombinant virus with ability to enter human cells has/has not emerged in the co-cultivations. This experiment is sought to be the experimental infection of *mus dunni* tail fibroblast cells (Methods 2.1) that are easily infectable by all MLVs and are widely used as an amplification assay for replication competent retrovirus. The assay traditionally allows *mus dunni* cells to grow for 21 days in the presence of putative virus to allow for sufficient amplification, but virus stocks are usually readily amplified from day 3 onwards.

We transfected *mus dunni* cells, with the addition of Polybrene 8ug/ml using supernatant from co-cultivations of passage 3 (days 2 and 10 post co-cultivation) and passage 4 (days 2 and 14 post co-cultivation) stem cells and a number of negative controls, ie human stem cell supernatant (growing as a monoculture) and CF1 cell supernatant alone. Supernatant from *mus dunni* cell transfections was collected at day 1, day 6 and day 14 post *mus dunni* cell transfection. The supernatants were then assayed by PERT, as described in Methods 2.5.7, in triplicate.

The results of the PERT assay (Figure 7.7) indicate that *mus dunni* tail fibroblast cells do not demonstrate any reproducible increase in RT activity when infected with the stem cell co-cultivation inoculates. Only two samples have shown a slight increase in RT activity, i.e. the stem co-cultivation sample from passage 3 day 10 (from the co-cultivation with Mitomycin treated CF1 cells) on the 14th day post *mus dunni* cell transfection and the sample from passage 4 day 2 (from the co-cultivation with non-Mitomycin treated CF1 cells) again on the 14th day post *mus dunni* cell transfection. As shown in Figure 7.7 these samples exhibit RT activity that peaks at 1.62E+03

picounits/µl and 2.79E+02 picounits/µl; both are below the limit of RT activity detection of 1E+04 picounits/µl for our optimised PERT assay. Triplicate samples however, do not measure equally increased RT values and show significant deviation.

As the positive control 22rv1 showed expected amplification peaks, this *mus dunni* assay appears to agree over the unlikely presence of a replication-competent propagating retrovirus in the stem cell co-cultivations. Importantly, the template and RT negative controls for the PERT assays all passed quality control for background noise and cannot account for the observed variability in the rest of the samples, even below the limit of detection.

Figure 7.7 PERT assay on supernatant from *mus dunni* cells transfected with stem cell supernatant from co-cultivations

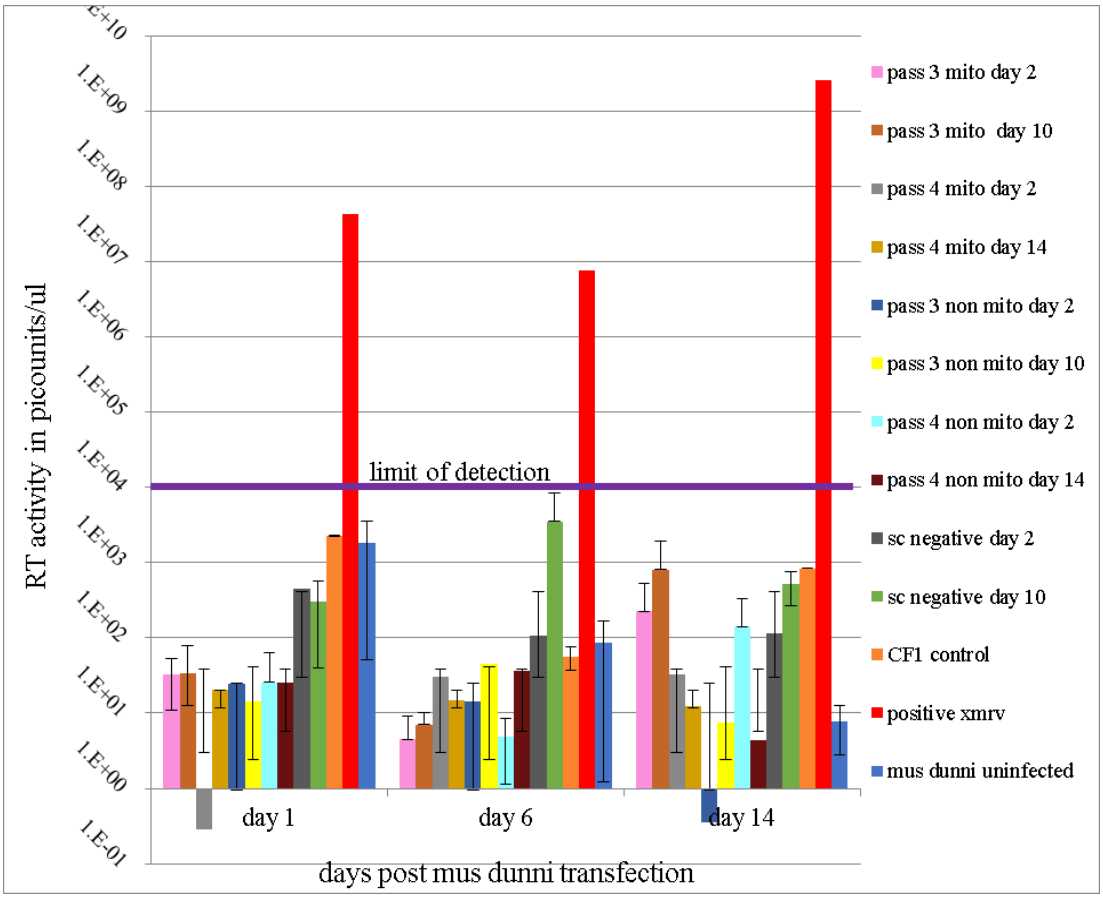


Figure 7.7 PERT assay on supernatant from *mus dunni* cells transfected with stem cell supernatant from co-cultivations

The figure demonstrates the RT activity measured in triplicate for various co-cultivation supernatants (stem cells with CF1 cells with or without Mitomycin) from passage 3 and 4. Negative controls comprised of monocultures of human stem cells grown alone in keratinocyte medium (taken from days 2 and 10) and uninfected *mus dunni* cells. PERT assays were performed on the infected *mus dunni* cells and controls on day 1,6 and 14.

7.8 CF1 cell-related MLV integration site in genomic DNA inferred from NGS of human stem cell co-cultivations

During analysis of the human stem cell and mouse CF1 cell co-cultivations with NGS, we observed that a large number of mouse related MLV integrations were being consistently described in negative controls such as in human stem cells alone and in the two supernatant infected stem cell populations (CF1 and XMRV). The mouse chromosome integrations discovered in these samples allowed us to infer that these were likely contaminations from reagents used in the amplification and/or sequencing process. These and their originating samples can be seen in Table 7.8.

Thus, as also shown in Table 7.8 we managed to rule out several of the mouse chromosomal integration sites amplified and sequenced as being due to contamination and we also pinpointed the location of several putative endogenous elements in the CF1 cell genome with varying levels of certainty as found in some but not all of the mouse containing cells. The sites were interrogated at the mouse reference genomic sequences for the full length virus (version mm9/ mm10). A variety of endogenous MLVs that were aligned with xenotropic and polytropic variants (Jern et al, 2007), were discovered in most of the integration sites, including the site described by Quinlan et al, 2011 in mouse chromosome 8; position 125,950,351 that is found in most of the samples including the negative controls. Indeed, for C57BL/6 that position is occupied by an *env*-lacking MLV of 3502bp total length with a 190bp insertion in the LTR (thus likely a MCF virus).

For the one site that we identified as the true possible site of the CF1-related MLV viral integration, ie position 125,712,672 of the mm9 version of mouse chromosome 8 in the C57BL/6 genome, there lies a full MLV that aligned perfectly with the CF1-related MLV consensus. When looking at version mm10, this position reiterated at 123,224,223 (AC163617.5). Interestingly, as the origin of CF1 mice is uncertain, we also interrogated other species reference gDNA apart from C57BL/6 using the Mouse Genomes Project variation viewer by the Sanger Institute and discovered eMLVs in

the AKR/J genome at position 123,430,577 (depth of coverage mean: 172x); the Balb_cj genome (34x), the CH3 genome (27x) and the NOD_ShiLtJ genome (19x) at the same positions +/- 100Kbp, missing only from the *mus spretus* genome. This is more close to the position described for emv-2 which is at ch8;123,426,292 (mm10) (see below paragraph).

All genomic copies in the reference sequences obtained, including the C57BL/6 genome position 123,224,223 do not bear the *pol* inactivating mutation discovered by Young et al, 2012a for emv-2. Additionally, when interrogating the variations in the e-MLV found in AKR/J we discovered that these are mainly A-G substitutions compared to emv-2 and are identical to the e-MLV discovered in this thesis in CF1 cells and to the emv-30 found in NOD_ShiLtJ mice (Triviai et al, 2014) and Prochazka 1991. Only two AA substitutions were found between CF1-related MLV (and emv-30) and the e-MLV in AKR/J in distal chromosome 8 and these were at positions: G3205A that codes for a tolerated mutation at R192K and G5601A, that is silent. Additionally the AKR/J e-MLV also lacks the characteristic LTR duplications and tandem repeats known for AKV-like viruses, making it a likely parent virus prior to AKV virus formation. We hypothesise this virus to be a likely parent for emv-30 and our CF1-MLV virus described in this thesis. A phylogenetic tree constructed for emv-30, the CF1-related MLV consensus sequence, emv-2 and the e-MLV discovered in AKR/J is shown in Figure 7.8.

Interestingly the polytropic virus amplifying in the direct sequencing of the integration site in CF1 cells was not seen throughout NGS for CF1 gDNA nor the co-cultivation samples. This may be due to a single or very few copies in the original genome of the CF1 cells.

Table 7.8 Mouse chromosome integration sequences in NGS samples

Mouse chromosome integration (position)	Mouse cell negative sample			Mouse cell and HSC co-cultivations				CF1 cells alone	Genomic result
	HSC alone	HSC infected with CF1	HSC infected with XMRV	pass 3 mito	pass 3 non mito	pass 4 mito	pass 4 non mito		
Ch1;138,614				x					**Xmv41/43**
Ch2;84,585	x								Contam.
Ch2;91,545	x								Contam.
Ch3;96,003	x	x							Contam.
Ch4;129,488				x	x				**Xmv8/9**
Ch5;114,987			x						Contam.
Ch8;26,552	x							x	Contam.
Ch8;86,715				x	x			x	**Xmv12**
Ch9;63,775								x	**Xmv15**
Ch8;125,712,672				x				x	**e-MLV**
Ch8;125,950,321	x	x	x			x	x	x	Contam.
Ch10;40,000				x	x				**Mpmv5**
Ch11;46,197	x	x							Contam.
Ch11;61,446	x	x						x	Contam.
Ch12;16,608			x						Contam.
Ch15;74,764			x						Contam.
Ch16;30,284	x							x	Contam.
Ch19;54,318	x	x	x			x	x	x	Contam.
ChX;136,538				x	x				**Pmv1**

Table 7.8 Mouse chromosome integration sequences in NGS samples

The Table shows the presence of repeated and unique mouse chromosome integration sequences for MLVs in the various samples tested in NGS, with or without the presence of mouse cells. This analysis further aided in the delineation of the exact and true integration site of the CF1-related MLV described in this thesis. In bold are putative MLVs possibly belonging to the CF1 endogenous reservoir, that however did not amplify in x/pMLV RT *gag* PCR.

Figure 7.8 Phylogenetic tree of MLV isolates in distal mouse ch8 (mm10)

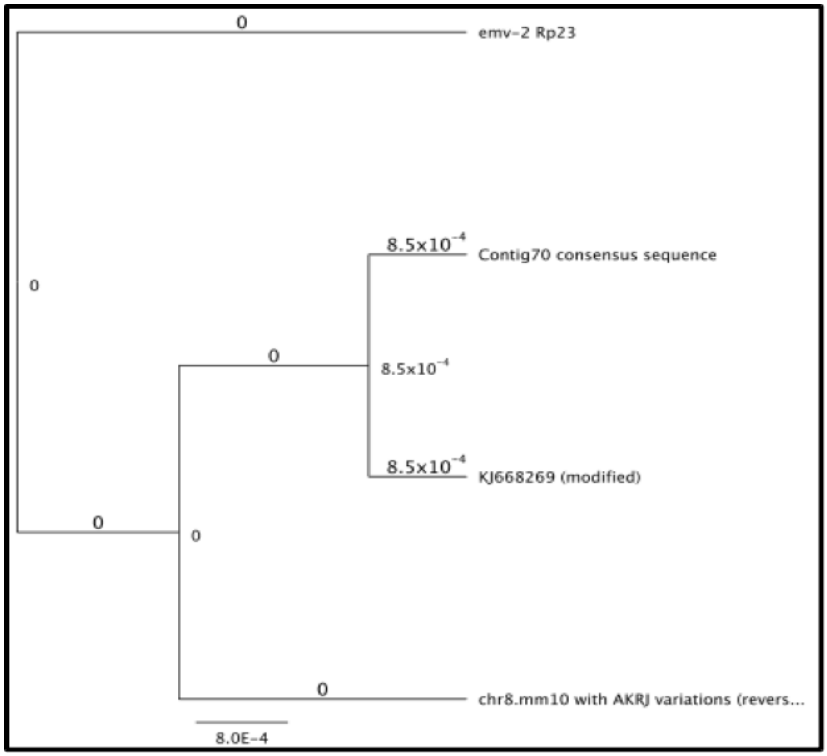


Figure 7.8 Phylogenetic tree of MLV isolates in distal mouse ch8 (mm10)

The tree shows the phylogenetic relationship as measured by base pair substitution ratio between the contig 70 consensus sequence (obtained from CF1 cells via sequencing in Chapter 6), KJ668269 (emv-30), emv-2 and the e-MLV discovered in distal chromosome 8 in the AKR/J structural variations database.

7.9 Discussion

In this chapter, we aimed to replicate the conditions used in cell transplantation practice, to cultivate human stem cells with the presence of murine feeder layers. The human stem cells derived from the human limbus at the corneoscleral junction have now been studied originally and post co-cultivation for the presence of immunostaining Anti- MLV Gag protein and the MLV *gag* and MLV messenger RNA *env* genes.

Indeed, MLV *gag* appears to increase over the course of the 3rd and 4th passage to reach quite high levels of MLV *gag* in co-cultivation supernatant. On the other hand, PERT has shown that there appears not to be a significant rise in RT activity associated with this MLV *gag* upsurge. Also, total cell RNA extracts do not appear to express MLV *gag* or mRNA *env* when controlled for mouse content. Hence, the upsurge in MLV *gag* in the co-cultivated cell supernatant could be due to the proliferative activity of the mouse feeders alone.

The characterisation of the MLV present in CF1 cells as ecotropic and the suggestion that it might be partially defective initially suggests that the infectivity potential for human cells is indeed low. However, MLV viruses have the ability to recombine between exogenous and endogenous proviruses that are found abundant in mice, thus allowing the emergence of a much greater phenotypic and genotypic variety and this seems to be a frequent case in immunocompromised strains (Young et al, 2012b). Indeed the phenomenon was first observed in mouse tumours, where recombination with endogenous viruses led to the birth of more pathogenic strains (Stoye et al, 1991).

To the above extent, we found no proof of the propagation of recombinant viruses in stem cell co-cultivations throughout 52 days. Indeed, we examined the co-cultivations for expressed virus with PERT, MLV *gag* PCR, spliced *env* PCR, *mus dunni* cell transfection and immunofluorescence. We also did not alter the clinical transplantation protocol to induce viral expression which may have produced different

results, of virological interest. The lack of an ecotropic receptor in human stem cells is also a point against the infection probability by the ecotropic CF1 –related MLV while endocytosis-mediated infection has only been shown dependent on the cell line used (Kubo et al, 2012). PH alteration or chemical stimulation within the additives used in stem cell culture, with unknown effects to viral propagation may also alter the stem cell culture conditions, although this was considered unlikely.

Interestingly, one publication examining the genomic stability of mouse stem cells, identified up to 41 clonal eMLV insertions via next generation sequencing and SNP analysis in mouse iPSCs derived from low passage Balb/C57BL hybrid mice after co-cultivation with one batch of CF1 feeder layers (Quinlan et al, 2011). When we compared the CF1 cell-related MLV sequence from the publication by Quinlan et al, 2011 we found that they pinpointed a region in the mouse reference genome mm9 in distal chromosome 8 qE1(125,949,348- 125,958,700bp). We have now shown that there exists no replication-competent e-MLV in that chromosomal position and this is a majorly defective remnant of an MCF virus copy. However, we believe that the true integration site of the CF1-related e-MLV virus is somewhat close i.e. in position 125,712,672 of the mm9 version of mouse chromosome 8 in the C57BL/6 genome. When looking at version mm10, this position reiterated at 123,224,223 (AC163617.5).

In this chapter we managed to show NGS detected clonal integration of XMRV in human stem cells and this served as a positive control for the depth such a fulminant infection produces. The observed low copy integration sequences are very minute to raise any concerns but have some discussion interest. Some indications of xMLVs or pMLVs presence is also noted in CF1 cells genomic DNA that are not seen in RNA sequencing (Chapter 6). This would probably necessitate a second next generation sequencing experiment for RNA full virus determination in the stem cell co-cultivations, not permitted by the length of this thesis. The absence of numerous integrations or an amplifiable RT signal however, indicate that the presence of a replication competent retrovirus that has gained the ability to be propagating is unlikely.

Chapter 8
DISCUSSION

In this thesis, an investigation of the transmission of viruses arising from endogenous ecotropic murine retroviruses (e-MLVs) in human stem cells was undertaken, using current in depth techniques and examining principles of recombinant virus formation in xenograft conditions. The possibility of the inadvertent rescue of an endogenous MLV with the ability to infect human primary stem cells and human stem cell lines is outlined as still ongoing in a variety of applications, clinical transplantation and mainly biomedical experimentation. Human stem cells are now being extensively tested to grow on murine feeders for the purposes of the creation of novel regenerative tissues, such as cartilage or bone, while attempts are being made to create artificial organs by using similar approaches, categorised as xenotransplantation.

The prototype methods used to test mouse fibroblast feeders for the presence or activation of endogenous murine retroviruses are outdated. The use of a dually labelled fluorescent probe (TaqMan) for the modernisation of the PERT assay has been attempted after review of previous studies. Moreover, a very large number of PERT assays conducted provided us with the experience and expertise to perform this assay on cell culture-derived supernatant and help from established researchers in the field (Dr Jeremy Garson) was also informative. The main difficulties arising in the effort to detect minute changes in enzymatic RT activity in MEF co-cultivations (Appendix 2) has been the likelihood of significant background noise arising from other material such as cellular polymerase enzymes and molecular reagents that are widely contaminated with mouse-derived sequences. The proof that Taq polymerases have endogenous RT-like activity was easily demonstrated in our experimental conditions and is also a novel unpublished finding.

Identifying positively amplified molecular targets in mouse-derived material is a daunting task that needs technical expertise, careful choice of controls and significant analytic skills. The presence of mouse-derived gDNA in laboratories through its use in monoclonal antibodies is widely documented (Erlwein et al, 2011) and we encountered

several instances where this was a confounding factor in our analysis. To this extent, we performed detailed cross-referencing with published sequences that although imperative, may have introduced a detection bias in our studies, which would have been absent in the molecular detection assays for any other known virus. The entirety of this work was conducted in two reference molecular laboratories with stringent protocols for clean PCR and additional measures were put in place to control for contamination, such as the use of mouse-DNA free extraction columns.

We experimented with a wide number of MLV stimulating agents and their combinations going back to the literature from the 1960s and 1970s (Chapter 3). Use of the Balb-K control for the formal induction experiments rendered our results valid, as this is currently the process recommended for such testing by the FDA Office of Biologicals.

One of the particular challenges arising from the co-cultivations of human and mouse stem cells was addressed by using molecular tools already known to control for such dual species targets; ie. the use of two reference house-keeping genes, one for human and one for mouse to derive human:mouse ratios was already described in Amit et al, 2005. The technical difficulty of growing cells and cell co-cultures on coverslips to produce confluent layers for confocal microscopy and TEM were issues encountered during our experimentation. We obtained a reporter human prostate cancer cell line, the DERSE-LnCap cells (Aloia et al, 2012) that was invaluable in that it is easily transfectable by supernatant and produces readily recognisable iGFP fluorescence.

The shift towards stem cell culture for regenerative surgery and the likelihood that these stem cell products may be considered in clinical transplantation, makes this present assessment imperative. Moreover, it is unknown what legislature exists outside the UK regarding xenotransplantation, especially in developing countries where research governance may be lacking. For example, human primary cells, such as limbal stem cells are now seldom grown on mouse feeder layers in the UK because non-mouse growth systems have been developed and adopted (Amit et al,2006; Yang et al,2013); examples are human feeders or amniotic membrane cultures (Yu et al, 2016) . Human stem cell lines, however, stored in the UK stem cell banks, are still being maintained and

propagated on MEFs (personal communication). It is unknown what the practice is for the rest of the world, but the cells mainly used in the UK are those of the J2 lineage and they have been assessed in this thesis. We can now offer an answer to the question of whether the J2-3T3 cell line is able to produce endogenous retroviruses when used in clinical transplantation protocols in ophthalmology, and this is highly unlikely (Chapter 3).

The discovery of an endogenous ecotropic MLV in one of the mouse feeder cell lines (CF1), used in stem cell experimentation is a relatively new find, as it is *not* documented in the insert that accompanies the cells. Although CF1 have a rather obscure origin, we have now tested at least 3 cell batches (Lot numbers: Appendix 11) from two different providers and have found that a unique MLV *gag* and *env* are expressed similarly in all with direct PCR and sequencing (Chapter 4). The full-length virus (GenBank Acc No MG525170) was characterised using novel NGS technology that shed significant light over the likelihood of variants, the presence or absence of oncogenes and/or finally provided with the confirmation of our molecular and imaging findings (Chapter 5). The capsid MHR related functional defect described is likely to have a significant impact on this endogenous e-MLV's ability to form virions. Similar but not the same mutations, in the neighbouring fv-1 region of the MHR have shown effects such as late virion maturation and loss of viral infectivity for deletion mutants (Cheslock et al, 2003) ; as the MHR is overall responsible for the early capsid protein assembly and dimer formation (Cairns et al, 2001). Interestingly, direct sequencing also demonstrated an endogenous polytropic isolate in CF1 cells that did not appear in viral expression experiments (Chapter 6 and 7).

The ability of NGS to provide with detailed and rather substantiated answers on a significant level of certainty, as offered by the number of reads that exceeds millions, is impressive (Schuster et al, 2008). Optimising this technique for our purposes has been challenging. We have been able to map a full length consensus of the virus (Chapter 6) and test co-cultivated cells for integration events caused by MLVs originating from feeder layers (Chapter 7). The literature failed to provide with a reference protocol for the optimisation of the particular conditions. The use of supernatant from cell culture has

not been previously documented for NGS but several exist for 454 pyrosequencing, i.e. the deep sequencing technology prior to Illumina (Marston et al, 2013) and others have successfully used NGS in various virus patient-derived samples.

Hence, we were confronted with several technical tasks that also introduced errors such as the inability to rid our samples completely of ribosomal RNA due to their prokaryotic content and detection of contaminating bacterial sequences thought to arise from the culture medium of cells, water (Laurence et al, 2014; Strong et al, 2014) or reagents (Salter et al, 2012). However, the existence of the positive control 22rv1 cell line that is a producer of XMRV aided in the standardisation of the results, as did the use of the internal RNA standard (MS2). By spending time re-iterating our analysis, we were able to search millions of sequences derived from both NGS experiments, ultimately being able to accurately describe fully the samples and their contents. The parsing of our RNA-seq reads at the PHE Bioinformatics department was also useful as this allowed for the relative quantification of variant reads using an in-house protocol. The use of the gene therapy tool, QuickMapper also proved informative for the characterisation of the integration sequences. Technical experience, however, dictates that such detailed sequencing analysis is done by elaborate operating system algorithms; where dedicated facilities can provide with automated analysis at a fraction of the time.

The discovery of a contaminating mouse sequences in NGS derived samples is not a surprise as, although not previously documented, falls in line with the documented presence of mouse DNA in laboratory instrumentation (Tuke et al, 2011).

The challenge of identifying true chimeric reads in the integration NGS presented several difficulties. The issue of chimeric reads needs to be distinguished as it is a likely to arise from ‘PCR jumping’; this is the effect of the random or chemical bond driven concatenation of two amplicons in one fragment, during the PCR amplification stage or the sequencing stage (Schloss et al, 2011). To avoid the effect of PCR amplification jumping, high fidelity proofreading Taq polymerase that controls for such errors was used (Bracho et al, 1996). For the sequencing reaction, the false amplification of chimeric reads is currently reported at 0.36- 1%. No particular solution exists; however,

we believe that in our samples this effect may have been less, as the sequences were additionally parsed for mouse viral-LTR sequences only; thus introducing a level of analytical discrimination. Some would suggest a pipeline approach that takes advantage of unpaired-end reads to infer the presence of the respective virus in the tested samples. (Wang et al, 2015). This would not be possible as our target virus is not DNA. The low number of reads associated with the integration sites discovered moreover, does indicate that these are likely to be within this <1% PCR artefact percentage. Although the confirmation arose from genomic DNA post co-cultivation that had not been subject to PCR, these were 5'-end LTR only sequences due to the design of our integration assay. Their very low numbers (2-3 copies) is, one may argue, sufficient for determination of risk vs no risk and their complete characterisation could indeed be the focus of further future experiments, not able to be covered in this thesis. Another approach would be that discussed in Forster et al, 2015 where they develop an algorithm to first map acquired reads to the human genome and then the viral isolates. This was not done in our study but was performed on the very low number of chimeric sequences after discovery. The fact that several of them (Figure 7.6.3) confirmed via PCR in the original genomic material is also important.

Another issue with handling large numbers of NGS sequences is the bias of false assignment that may be present in indexes allocation, due to cross contamination or cluster misidentification. This has been reported to be present in raw data arising from NGS multiplexed experiments in as much as 0.3% and is sought to be eliminated if a stringent quality filter is applied (minimum 15) (Illumina). The use of an internal spiking control also may help eliminate this error, albeit only at the RNA level in our experiments. Contamination at the library preparation stage is also a very common occurrence that is attributed to technical error at the preparation of sample pools.

In our experiments we discovered 4 integration sequences that were incorrectly assigned to different sample indexes, originating initially from XMRV (data not shown). The raw data was re-examined and the correct indexes were discovered. The issue, therefore, of incorrect cluster allocation by the sequencer may also need some additional attention. Several additional algorithms have been designed that are able to weigh alignments

derived from NGS and are sought to control for index contamination or sequencer misallocation however these were not used.

More importantly, however, for viral full-length consensus building and variant support RNA-seq has a limitation that corresponds to the errors in reads generated and errors in reads assembled that correspond to repetitive regions of the reference sequence. We did not account for this error in our experiment in Chapter 5. This means that insertion or deletions shown in variant base calling were also likely to be confounded by sequencing errors. These are all limitations that could be sought in a larger scale experiment, while the area of new virus discovery and characterisation through NGS is still being developed.

This project has contributed significantly to the area of endogenous MLV virus presence, detection and transmission for human primary stem cells, of interest in xenotransplantation practice. I have demonstrated that although xenotropic MLVs such as XMRV are able to infect human limbal stem cells and integrate into genomic DNA in a clonal fashion, this is not likely to be the case for established ecotropic MLVs arising from cell culture. Interestingly, the detection and confirmation of a minute number of integrated MLV sequences in human limbal stem cell gDNA further strengthens the principle; ie that only replication-competent retroviruses are able to productively infect cells with a likelihood of an *in vivo* effect. The integration sites discovered are always likely to be infection events at a single cell level from recombinant viruses arising in established ecotropic MLV cell infection (corroborated by the single unconfirmed indication that a polytropic proviral copy is also present in CF1 gDNA). To this extent it is important that we were able to show that the integration sequences were present in the samples prior to the introduction of the NGS sequencing reaction and library preparation. However, these were only confirmed on the 5'-end and extensive experiments would be further needed to demonstrate single- cell integrations in the original genomic material.

The CF1-related MLV is an ecotropic MLV that is similar to the described emv-30 from NOD_ShiLtJ mice; that also bear a polytropic copy that was not found expressed in CF1 cells via RNA-seq or conventional PCR for *gag* or *env*. All the above may indeed be an

indication that there exists a common progenitor between the two murine strains that is previously undocumented and the origins of the CF1 cells are not CF1 mice but NOD.

Further research could focus on the demonstration of the true recombinant virus genesis in a prolonged human stem cell co-cultivation environment; perhaps even with the use of a three-dimensional culture system that is likely to be used in tissue and organ generation in the near future (Johnson et al, 2007). Further experiments should be able to demonstrate the clonal and predominant infection of human stem cells by xenotropic recombinants other than XMRV, arising in cell culture *in vitro*. Our demonstration that XMRV infects and integrates in human limbal primary stem cells via supernatant alone, adds to the proof that the transplantation of human cells grown on mouse feeder layers containing xenotropic endogenous viruses capable of activation should now be completely abolished.

Further experiments could also focus on the effect these unique MLV integrations may have on the differentiation and/or function of the human stem cells; where integration may be likely to cause from generic ie *a propos* of *de novo* methylation processes to selective insertional mutagenesis relevant to the gene(s) affected. Apart from the infectivity issues for human stem cells that have been addressed in this thesis with NGS, there is extensive proof that viral induction may affect other important cell biology experiments such as TLR signalling via Nf-KB activation (Hiscott et al, 2006), or G-coupling receptor studies (Vaughan et al, 2012). A wide number of hypotheses stem from such experiments, where awareness and consideration of an expressed viral element is important.

REFERENCES

Aaronson, S. A., Hartley, J. W., & Todaro, G. J. (1969). Mouse leukaemia virus: "spontaneous" release by mouse embryo cells after long-term in vitro cultivation. Proceedings of the National Academy of Sciences of the United States of America, 64(1), 87–94.

Aaronson, S. A., Todaro, G. J., & Scolnick, E. M. (1971). Induction of murine C-type viruses from clonal lines of virus-free BALB-3T3 cells. Science (New York, N.Y.), 174(4005), 157–159.

Accola, M. A., Höglund, S., & Göttlinger, H. G. (1998). A putative alpha-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. Journal of Virology, 72(3), 2072–2078.

Alam, S. L., Wills, N. M., Ingram, J. A., Atkins, J. F., & Gesteland, R. F. (1999). Structural studies of the RNA pseudoknot required for readthrough of the gag-termination codon of murine leukaemia virus. Journal of Molecular Biology, 288(5), 837–852.

Alberts, B. (2011). Retraction. Science (New York, N.Y.), 334(6063), 1636.

Aloia, A. L., Duffy, L., Pak, V., Lee, K. E., Sanchez-Martinez, S., Derse, D., … Rein, A. (2012). A reporter system for replication-competent gammaretroviruses: the inGluc-MLV-DERSE assay. Gene Therapy. 2013 Feb;20(2):169-76

Alter, H. J., Mikovits, J. A., Switzer, W. M., Ruscetti, F. W., Lo, S.-C., Klimas, N., … Lipkin, W. I. (2012). A multicenter blinded analysis indicates no association between chronic fatigue syndrome/myalgic encephalomyelitis and either xenotropic murine leukaemia virus-related virus or polytropic murine leukaemia virus. mBio, 3(5).

Amit, M., Winkler, M. E., Menke, S., Brüning, E., Büscher, K., Denner, J., … Martin, U. (2005). No evidence for infection of human embryonic stem cells by feeder cell-derived murine leukaemia viruses. Stem Cells (Dayton, Ohio), 23(6), 761–771.

Amit M, Itskovitz-Eldor J. (2006) Feeder-free culture of human embryonic stem cells. Methods Enzymol. 420:37-49.

Andersen, K. B., & Nexø, B. A. (1983). Entry of murine retrovirus into mouse fibroblasts. Virology, 125(1), 85–98.

Ashe, M. P., Pearson, L. H., & Proudfoot, N. J. (1997). The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. The EMBO Journal, 16(18), 5752–5763.

Baldo, A. M., & McClure, M. A. (1999). Evolution and horizontal transfer of dUTPase-encoding genes in viruses and their hosts. Journal of Virology, 73(9), 7710–7721.

Barnett, A. L., & Cunningham, J. M. (2001). Receptor binding transforms the surface subunit of the mammalian C-type retrovirus envelope protein from an inhibitor to an activator of fusion. Journal of Virology, 75(19), 9096–9105.

Battini, J. L., Kayman, S. C., Pinter, A., Heard, J. M., & Danos, O. (1994). Role of N-linked glycosylation in the activity of the Friend murine leukaemia virus SU protein receptor-binding domain. Virology, 202(1), 496–499.

Baylis, O., Figueiredo, F., Henein, C., Lako, M., & Ahmad, S. (2011). 13 years of cultured limbal epithelial cell therapy: a review of the outcomes. Journal of Cellular Biochemistry, 112(4), 993–1002.

Bender, W., Chien, Y. H., Chattopadhyay, S., Vogt, P. K., Gardner, M. B., & Davidson, N. (1978). High-molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures. Journal of Virology, 25(3), 888–896.

Bernhard, W. (1958). Electron microscopy of tumor cells and tumor viruses; a review. Cancer Research, 18(5), 491–509.

Bisbal, C., & Salehzada, T. (2008). [RNase L, a crucial mediator of innate immunity and other cell functions]. Médecine Sciences: M/S, 24(10), 859–864.

Blackard JT, Cohen DE, Mayer KH. (2002) Human immunodeficiency virus superinfection and recombination: current state of knowledge and potential clinical consequences. Clin Infect Dis. Apr 15;34(8):1108-14.

Blacklow, S. C., Lu, M., & Kim, P. S. (1995). A trimeric subdomain of the simian immunodeficiency virus envelope glycoprotein. Biochemistry, 34(46), 14955–14962.

Briggs, J. A. G., Simon, M. N., Gross, I., Kräusslich, H.-G., Fuller, S. D., Vogt, V. M., & Johnson, M. C. (2004). The stoichiometry of Gag protein in HIV-1. Nature Structural & Molecular Biology, 11(7), 672–675.

Brooks, J., Lycett-Lambert, K., Caminiti, K., Merks, H., McMillan, R., & Sandstrom, P. (2011). No evidence of cross-species transmission of mouse retroviruses to animal workers exposed to mice. Transfusion.Feb;52(2):317-25.

Buck, B. M., & Bather, R. (1969). Requirements for RNA and DNA Synthesis in the Transformation of Mouse Embryo Cells by Murine Sarcoma Virus (Moloney). Journal of General Virology, 4(3), 457–460.

Burton , D.R., Desrosiers, R.C., Doms, R.W., Koff, W.C., Kwong, P.D., Moore, J.P., Nabel, G.J., Sodroski J., Wilson, I.A., and Wyatt, R.T. (2004) HIV vaccine design and the neutralising antibody problem. Nat. Immunol. 5,233-236

Caffrey, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Gronenborn, A. M., & Clore, G. M. (1998). 3D NMR experiments for measuring 15N relaxation data of large

proteins: application to the 44 kDa ectodomain of SIV gp41. Journal of Magnetic Resonance (San Diego, Calif.: 1997), 135(2), 368–372.

Cairns, T. M., & Craven, R. C. (2001). Viral DNA synthesis defects in assembly-competent Rous sarcoma virus CA mutants. Journal of Virology, 75(1), 242–250.

Carteau, S., Gorelick, R. J., & Bushman, F. D. (1999). Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. Journal of Virology, 73(8), 6670–6679.

Chahwan, R., Wontakal, S. N., & Roa, S. (2010). Crosstalk between genetic and epigenetic information through cytosine deamination. Trends in Genetics: TIG, 26(10), 443–448.

Chandler, L. A., & Jones, P. A. (1988). Hypomethylation of DNA in the regulation of gene expression. Developmental Biology (New York, N.Y.: 1985), 5, 335–349.

Chen, E. C., Miller, S. A., DeRisi, J. L., & Chiu, C. Y. (2011). Using a pan-viral microarray assay (Virochip) to screen clinical samples for viral pathogens. Journal of Visualized Experiments: JoVE, (50).

Cherry, S. R., Biniszkiewicz, D., van Parijs, L., Baltimore, D., & Jaenisch, R. (2000). Retroviral expression in embryonic stem cells and hematopoietic stem cells. Molecular and Cellular Biology, 20(20), 7419–7426.

Cheslock, S. R., Poon, D. T. K., Fu, W., Rhodes, T. D., Henderson, L. E., Nagashima, K., … Hu, W.-S. (2003). Charged Assembly Helix Motif in Murine Leukemia Virus Capsid: an Important Region for Virus Assembly and Particle Size Determination. Journal of Virology, 77(12), 7058–7066.

Chow S.A.,Vincent, K.A., Ellison, V., and Brown, P.O. (1992). Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. Science 255,723-726

Cingöz, O., Paprotka, T., Delviks-Frankenberry, K. A., Wildt, S., Hu, W.-S., Pathak, V. K., & Coffin, J. M. (2011). Characterization, Mapping and Distribution of the Two XMRV Parental Proviruses. Journal of Virology. Jan;86(1):328-38

Coffin, J. M. (1979) Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. J. Gen. Virol. 42, 1–26.

Coffin J.M. , Hughes S.H. and Varmus H.E. (1997) ; Retroviruses. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Coffin J.M. , Fan H. (2016); The discovery of reverse transcriptase. Annu. Rev. Virol., 3, pp 29-51.

Cosset, F. L., Takeuchi, Y., Battini, J. L., Weiss, R. A., & Collins, M. K. (1995). High-titer packaging cells producing recombinant retroviruses resistant to human serum. J. Virol., 69(12), 7430–7436.

Cullen, B. R., Lomedico, P. T., & Ju, G. (1984). Transcriptional interference in avian retroviruses--implications for the promoter insertion model of leukaemogenesis. Nature, 307(5948), 241–245.

Cullen, B.R., (1992) Mechanism of action of regulatory proteins encoded by complex retroviruses, Microbiol. Rev. 56, 375-394

Daya, S. M., Watson, A., Sharpe, J. R., Giledi, O., Rowe, A., Martin, R., & James, S. E. (2005). Outcomes and DNA analysis of ex vivo expanded stem cell allograft for ocular surface reconstruction. Ophthalmology, 112(3), 470–477.

Denner, J., & Tonjes, R. R. (2012). Infection Barriers to Successful Xenotransplantation Focusing on Porcine Endogenous Retroviruses. Clinical Microbiology Reviews, 25(2), 318–343.

D'Souza, V., & Summers, M. F. (2004). Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. Nature, 431(7008), 586–590.

D'Souza, V., & Summers, M. F. (2005). How retroviruses select their genomes. Nature Reviews. Microbiology, 3(8), 643–655.

Davey, R.A., Hamson, C.A., Healey, J.J. and Cunnigham J.M. (1997) In vitro binding of purified murine ecotropic retrovirus envelope surface protein to its receptor, MCAT-1. J.Virol. 71, 8096-8102

Dua, H. S., & Azuara-Blanco, A. (2000). Limbal stem cells of the corneal epithelium. Survey of Ophthalmology, 44(5), 415–425.

Dupressoir, A., Vernochet, C., Harper, F., Guégan, J., Dessen, P., Pierron, G., & Heidmann, T. (2011). A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. Proceedings of the National Academy of Sciences of the United States of America, 108(46), E1164–1173.

Erlwein, O., Kaye, S., McClure, M. O., Weber, J., Wills, G., Collier, D., … Cleare, A. (2010). Failure to detect the novel retrovirus XMRV in chronic fatigue syndrome. PloS One, 5(1), e8519.

Erlwein, O., Robinson, M. J., Kaye, S., Wills, G., Izui, S., Wessely, S., … McClure, M. O. (2011). Investigation into the presence of and serological response to XMRV in CFS patients. PloS One, 6(3), e17592.

Evans, L. H., Alamgir, A. S. M., Owens, N., Weber, N., Virtaneva, K., Barbian, K., … Rosenke, K. (2009). Mobilization of endogenous retroviruses in mice after infection with an exogenous retrovirus. Journal of Virology, 83(6), 2429–2435. h

Felsenstein, K. M., & Goff, S. P. (1988). Expression of the gag-pol fusion protein of Moloney murine leukaemia virus without gag protein does not induce virion formation or proteolytic processing. Journal of Virology, 62(6), 2179–2182.

Ferrarone, J., Knoper, R. C., Li, R., & Kozak, C. A. (2012). Second site mutation in the virus envelope expands the host range of a cytopathic variant of Moloney murine leukaemia virus. Virology, 433(1), 7–11.

Flotho, C., Claus, R., Batz, C., Schneider, M., Sandrock, I., Ihde, S., … Lübbert, M. (2009). The DNA methyltransferase inhibitors azacitidine, decitabine and zebularine exert differential effects on cancer gene expression in acute myeloid leukaemia cells. Leukemia, 23(6), 1019–1028.

Förster, F., Medalia, O., Zauberman, N., Baumeister, W., & Fass, D. (2005). Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. Proceedings of the National Academy of Sciences of the United States of America, 102(13), 4729–4734.

Forster M, Szymczak S, Ellinghaus D, Hemmrich G, Rühlemann M, Kraemer L, Mucha S, Wienbrandt L, Stanulla M; UFO Sequencing Consortium within I-BFM Study Group, Franke A. (2015) Vy-PER: eliminating false positive detection of virus integration events in next generation sequencing data. Sci Rep. 13;5:11534.

Garson, J. A., Kellam, P., & Towers, G. J. (2011). Analysis of XMRV integration sites from human prostate cancer tissues suggests PCR contamination rather than genuine human infection. Retrovirology, 8, 13.

Gimbel, W., Schmidt, J., Brack-Werner, R., Luz, A., Strauss, P. G., Erfle, V., & Werner, T. (1996). Molecular and pathogenic characterization of the RFB osteoma virus: lack of oncogene and induction of osteoma, osteopetrosis, and lymphoma. Virology, 224(2), 533–538.

Glinsky, G. V. (2015). Viruses, stemness, embryogenesis, and cancer: a miracle leap toward molecular definition of novel oncotargets for therapy-resistant malignant tumors? Oncoscience, 2(9), 751–754.

Goff, S. P. (2007). Retroviridae: the retroviruses and their replication, p. 1999-2069. In D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus (ed.), Fields virology, 5th ed. Lippincott Williams & Wilkins, Philadelphia, PA

Gopalakrishnan, V., Peliska, J. A., & Benkovic, S. J. (1992). Human immunodeficiency virus type 1 reverse transcriptase: spatial and temporal relationship between the polymerase and RNase H activities. Proceedings of the National Academy of Sciences of the United States of America, 89(22), 10763–10767.

Gray, K. D., & Roth, M. J. (1993). Mutational analysis of the envelope gene of Moloney murine leukaemia virus. Journal of Virology, 67(6), 3489–3496.

Green, H. (2008). The birth of therapy with cultured cells. BioEssays, 30(9), 897–903.

Griffiths, D. J. (2001). Endogenous retroviruses in the human genome sequence. Genome Biology, 2(6), reviews1017.1–reviews1017.5.

Groom, H. C. T., Boucherit, V. C., Makinson, K., Randal, E., Baptista, S., Hagan, S., … Bishop, K. N. (2010). Absence of xenotropic murine leukaemia virus-related virus in UK patients with chronic fatigue syndrome. Retrovirology, 7, 10.

Groom, H. C. T., Yap, M. W., Galão, R. P., Neil, S. J. D., & Bishop, K. N. (2010). Susceptibility of xenotropic murine leukaemia virus-related virus (XMRV) to retroviral restriction factors. Proceedings of the National Academy of Sciences of the United States of America, 107(11), 5166–5171.

Grueterich, M., Espana, E. M., Touhami, A., Ti, S.-E., & Tseng, S. C. G. (2002). Phenotypic study of a case with successful transplantation of ex vivo expanded human limbal epithelium for unilateral total limbal stem cell deficiency. Ophthalmology, 109(8), 1547–1552.

Hagemann, S., Heil, O., Lyko, F., & Brueckner, B. (2011). Azacytidine and decitabine induce gene-specific and non-random DNA demethylation in human cancer cell lines. PloS One, 6(3), e17388.

Hare, S., Gupta, S. S., Valkov, E., Engelman, A., & Cherepanov, P. (2010). Retroviral intasome assembly and inhibition of DNA strand transfer. Nature, 464(7286), 232–236.

Hartley, J. W., & Rowe, W. P. (1976). Naturally occurring murine leukaemia viruses in wild mice: characterization of a new “amphotropic” class. Journal of Virology, 19(1), 19–25.

Hartley, J. W., Rowe, W. P., & Huebner, R. J. (1970). Host-range restrictions of murine leukaemia viruses in mouse embryo cell cultures. Journal of Virology, 5(2), 221–225.

Heard, J. M., & Danos, O. (1991). An amino-terminal fragment of the Friend murine leukaemia virus envelope glycoprotein binds the ecotropic receptor. Journal of Virology, 65(8), 4026–4032.

Hiscott J, Nguyen TL, Arguello M, Nakhaei P, Paz S. (2006) Manipulation of the nuclear factor-kappaB pathway and the innate immune response by viruses. Oncogene. Oct 30;25(51):6844-67.

Hizi A., Herzig E. (2015) dUTPase: the frequently overlooked enzyme encoded by many retroviruses. Retrovirology Aug 12;12:70

Horoszewicz, J. S., Leong, S. S., Kawinski, E., Karr, J. P., Rosenthal, H., Chu, T. M., … Murphy, G. P. (1983). LNCaP model of human prostatic carcinoma. Cancer Research, 43(4), 1809–1818.

Huber, H. E., & Richardson, C. C. (1990). Processing of the primer for plus strand DNA synthesis by human immunodeficiency virus 1 reverse transcriptase. The Journal of Biological Chemistry, 265(18), 10565–10573.

Hué, S., Gray, E. R., Gall, A., Katzourakis, A., Tan, C. P., Houldcroft, C. J., … Towers, G. J. (2010). Disease-associated XMRV sequences are consistent with laboratory contamination. Retrovirology, 7(1), 111.

Hu, W. S., & Temin, H. M. (1990). Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic

recombination. Proceedings of the National Academy of Sciences of the United States of America, 87(4), 1556–1560.

Hughes SH (2015) Reverse Transciption of Retroviruses and LTR Retrotransposons; Microbiol Spectr. Apr; 3(2)

Hwang CK, Svarovskaia ES, Pathak VK. (2001) Dynamic copy choice: steady state between murine leukaemia virus polymerase and polymerase-dependent RNase H activity determines frequency of in vivo template switching. Proc Natl Acad Sci U S A. 9;98(21):12209-14.

Innes, C. L., Smith, P. B., Langenbach, R., Tindall, K. R., & Boone, L. R. (1990). Cationic liposomes (Lipofectin) mediate retroviral infection in the absence of specific receptors. Journal of Virology, 64(2), 957–961.

Ishida, T., Hamano, A., Koiwa, T., & Watanabe, T. (2006). 5' long terminal repeat (LTR)-selective methylation of latently infected HIV-1 provirus that is demethylated by reactivation signals. Retrovirology, 3(1), 1–7.

Jacks, T., & Varmus, H. E. (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science (New York, N.Y.), 230(4731), 1237–1242.

Jainchill, J. L., Aaronson, S. A., & Todaro, G. J. (1969). Murine sarcoma and leukaemia viruses: assay using clonal lines of contact-inhibited mouse cells. Journal of Virology, 4(5), 549–553.

Jern, P., Stoye, J. P., & Coffin, J. M. (2007). Role of APOBEC3 in Genetic Diversity among Endogenous Murine Leukemia Viruses. PLoS Genetics, 3(10), e183.

Johnen, S., Wickert, L., Meier, M., Salz, A. K., Walter, P., & Thumann, G. (2011). Presence of xenogenic mouse RNA in RPE and IPE cells cultured on mitotically inhibited 3T3 fibroblasts. Investigative Ophthalmology & Visual Science, 52(5), 2817–2824.

Johnson, K. R., Leight, J. L., & Weaver, V. M. (2007). Demystifying the Effects of a Three−Dimensional Microenvironment in Tissue Morphogenesis. In Methods in Cell Biology (Vol. 83, pp. 547–583). Elsevier.

Kakisi, O. K., Robinson, M. J., Tettmar, K. I., & Tedder, R. S. (2013). The rise and fall of XMRV. Transfusion Medicine (Oxford, England), 23(3), 142–151.

Katoh I, Yasunaga T, Ikawa Y, Yoshinaka Y. (1987) Inhibition of retroviral protease activity by an aspartyl proteinase inhibitor . Nature Oct 15-21;329(6140):654-6

Katz R.A., Greger J.G., Skalka A.M. (2005) Effects of cell cycle status on early events in tr

Kawai, S., & Nishizawa, M. (1984). New procedure for DNA transfection with polycation and dimethyl sulfoxide. Molecular and Cellular Biology, 4(6), 1172–1174.

Khan, A. S., Muller, J., & Sears, J. F. (2001). Early detection of endogenous retroviruses in chemically induced mouse cells. Virus Research, 79(1-2), 39–45.

Kim, J. K., Palaniappan, C., Wu, W., Fay, P. J. & Bambara, R. A. (1997) Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1. J. Biol. Chem. 272, 16769–16777.

Kim, S., Rusmevichientong, A., Dong, B., Remenyi, R., Silverman, R. H., & Chow, S. A. (2010). Fidelity of Target Site Duplication and Sequence Preference during Integration of Xenotropic Murine Leukemia Virus-Related Virus. PLoS ONE, 5(4), e10255.

Klement, V., Rowe, W. P., Hartley, J. W., & Pugh, W. E. (1969). Mixed culture cytopathogenicity: a new test for growth of murine leukaemia viruses in tissue culture. Proceedings of the National Academy of Sciences of the United States of America, 63(3), 753–758.

Knoper RC, Ferrarone J, Yan Y, Lafont BA, Kozak CA (2009) Removal of either N-glycan site from the envelope receptor binding domain of Moloney and Friend but not AKV mouse ecotropic gammaretroviruses alters receptor usage. Virology. Sep 1;391(2):232-9.

Knouf, E. C., Metzger, M. J., Mitchell, P. S., Arroyo, J. D., Chevillet, J. R., Tewari, M., & Miller, A. D. (2009). Multiple integrated copies and high-level production of the human retrovirus XMRV (xenotropic murine leukaemia virus-related virus) from 22Rv1 prostate carcinoma cells. Journal of Virology, 83(14), 7353–7356.

Knox, K., Carrigan, D., Simmons, G., Teque, F., Zhou, Y., Hackett, J., … Levy, J. A. (2011). No evidence of murine-like gammaretroviruses in CFS patients previously identified as XMRV-infected. Science (New York, N.Y.), 333(6038), 94–97.

Koch, W., Hunsmann, G., & Friedrich, R. (1983). Nucleotide sequence of the envelope gene of Friend murine leukaemia virus. Journal of Virology, 45(1), 1–9.

Kozak, C. A. (2010). The mouse “xenotropic” gammaretroviruses and their XPR1 receptor. Retrovirology, 7, 101.

Kozak, C. A. (2011). Naturally Occurring Polymorphisms of the Mouse Gammaretrovirus Receptors CAT-1 and XPR1 Alter Virus Tropism and Pathogenicity. Advances in Virology, 2011, 975801.

Kozak, C. A., & Chakraborti, A. (1996). Single amino acid changes in the murine leukaemia virus capsid protein gene define the target of Fv1 resistance. Virology, 225(2), 300–305.

Kräusslich, H. G., Fäcke, M., Heuser, A. M., Konvalinka, J., & Zentgraf, H. (1995). The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. Journal of Virology, 69(6), 3407–3419.

Kubo, Y., Hayashi, H., Matsuyama, T., Sato, H., & Yamamoto, N. (2012). Retrovirus Entry by Endocytosis and Cathepsin Proteases. Advances in Virology, 640894.

Kurth, R., & Bannert, N. (2010a). Beneficial and detrimental effects of human endogenous retroviruses. International Journal of Cancer, 126(2), 306–314.

Kurth, R., & Bannert, N. (2010b) Retroviruses: Molecular Biology, Genomics and Pathogenesis. Caister Academic Press

Lanciault, C., & Champoux, J. J. (2006). Pausing during Reverse Transcription Increases the Rate of Retroviral Recombination. Journal of Virology, 80(5), 2483–2494.

Lander, M. R., & Chattopadhyay, S. K. (1984). A Mus dunni cell line that lacks sequences closely related to endogenous murine leukaemia viruses and can be infected by ectropic, amphotropic, xenotropic, and mink cell focus-forming viruses. Journal of Virology, 52(2), 695–698.

Lasneret, J., Canivet, M., Hojman-Montes de Oca, F., Tobaly, J., Emanoil-Ravicovitch, R., & Peries, J. (1983). Activation of intracistemal a particles by 5-azacytidine in mouse Ki-BALB cell line. Virology, 128(2), 485–489.

Laurence, M., Hatzis, C., & Brash, D. E. (2014). Common Contaminants in Next-Generation Sequencing That Hinder Discovery of Low-Abundance Microbes. PLoS ONE, 9(5), e97876.

Lavillette, D., & Kabat, D. (2004). Porcine endogenous retroviruses infect cells lacking cognate receptors by an alternative pathway: implications for retrovirus evolution and xenotransplantation. Journal of Virology, 78(16), 8868–8877.

Lavillette, D., Maurice, M., Roche, C., Russell, S. J., Sitbon, M., & Cosset, F. L. (1998). A proline-rich motif downstream of the receptor binding domain modulates conformation and fusogenicity of murine retroviral envelopes. Journal of Virology, 72(12), 9955–9965.

Leib-Mösch, C., Schmidt, J., Etzerodt, M., Pedersen, F. S., Hehlmann, R., & Erfle, V. (1986). Oncogenic retrovirus from spontaneous murine osteomas. II. Molecular cloning and genomic characterization. Virology, 150(1), 96–105.

Lewis P, Hensel M, Emerman M. (1992) Human immunodeficiency virus infection of cells arrested in the cell cycle. EMBO J Aug;11(8):3053-8.

Lim, K. (2012). Retroviral infection of hES cells produces random-like integration patterns. Molecules and Cells, 33(5), 525–531.

Li, W., Sabater, A. L., Chen, Y.-T., Hayashida, Y., Chen, S.-Y., He, H., & Tseng, S. C. G. (2007). A novel method of isolation, preservation, and expansion of human corneal endothelial cells. Investigative Ophthalmology & Visual Science, 48(2), 614–620.

Lloyd, A., White, P., Wessely, S., Sharpe, M., & Buchwald, D. (2010). Comment on "Detection of an infectious retrovirus, XMRV, in blood cells of patients with chronic fatigue syndrome." Science (New York, N.Y.), 328(5980), 825; author reply 825.

Loiler, S. A., DiFronzo, N. L., & Holland, C. A. (1997). Gene transfer to human cells using retrovirus vectors produced by a new polytropic packaging cell line. Journal of Virology, 71(6), 4825–4828.

Lombardi, V. C., Ruscetti, F. W., Das Gupta, J., Pfost, M. A., Hagen, K. S., Peterson, D. L., … Mikovits, J. A. (2009). Detection of an infectious retrovirus, XMRV, in blood cells of patients with chronic fatigue syndrome. Science (New York, N.Y.), 326(5952), 585–589.

Lo, S.-C., Pripuzova, N., Li, B., Komaroff, A. L., Hung, G.-C., Wang, R., & Alter, H. J. (2010). Detection of MLV-related virus gene sequences in blood of patients with chronic fatigue syndrome and healthy blood donors. Proceedings of the National Academy of Sciences of the United States of America, 107(36), 15874–15879.

Lusso, P., di Marzo Veronese, F., Ensoli, B., Franchini, G., Jemma, C., DeRocco, S., … Gallo, R. (1990). Expanded HIV-1 cellular tropism by phenotypic mixing with murine endogenous retroviruses. Science, 247(4944), 848 –852.

Luo GX, Taylor J. (1990) Template switching by reverse transcriptase during DNA synthesis. J Virol. Sep;64(9):4321-8.

Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. (2006) Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS genetics. 2(1):e2

Marcheschi, R. J., Staple, D. W., & Butcher, S. E. (2007). Programmed ribosomal frameshifting in SIV is induced by a highly structured RNA stem-loop. Journal of Molecular Biology, 373(3), 652–663.

Margalith, M., Thornton, H., Narconis, R., Pinkerton, H., & Green, M. (1975). Studies on virus induction by 5-bromodeoxyuridine in nonproducer murine sarcoma virus-transformed 3T3 cells. Virology, 65(1), 27–39.

Marriott, S. J., & Semmes, O. J. (2005). Impact of HTLV-I Tax on cell cycle progression and the cellular DNA damage repair response. Oncogene, 24(39), 5986–5995.

Marston, D. A., McElhinney, L. M., Ellis, R. J., Horton, D. L., Wise, E. L., Leech, S. L., … Fooks, A. R. (2013). Next generation sequencing of viral RNA genomes. BMC Genomics, 14, 444.

Martin J, Kabat P, Tristem M (2003) : Cospeciation and horizontal transmission in the murine leukaemia-related retroviruses. Tangled trees: phylogenies, cospeciation, and coevolution. Ed Page RDM. 2002, Chicago and London, University of Chicago Press, 174-194

McClure, M. O., Sommerfelt, M. A., Marsh, M., & Weiss, R. A. (1990). The pH independence of mammalian retrovirus infection. The Journal of General Virology, 71 (Pt 4), 767–773.

Mendoza, R., Vaughan, A. E., & Miller, A. D. (2011). The Left Half of the XMRV Retrovirus Is Present in an Endogenous Retrovirus of NIH/3T3 Swiss Mouse Cells. J. Virol., 85(17), 9247–9248.

Metzger, M. J., Holguin, C. J., Mendoza, R., & Miller, A. D. (2010). The prostate cancer-associated human retrovirus XMRV lacks direct transforming activity but can induce low rates of transformation in cultured cells. Journal of Virology, 84(4), 1874–1880.

Mikkers, H., Allen, J., Knipscheer, P., Romeijn, L., Hart, A., Vink, E., … Romeyn, L. (2002). High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. Nature Genetics, 32(1), 153–159.

Moloney J.B. (1960) Properties of a leukaemia virus. Natl.Cancer Inst. Monogr. 4, 7-37

Morgan, R. A., Nussbaum, O., Muenchau, D. D., Shu, L., Couture, L., & Anderson, W. F. (1993). Analysis of the functional and host range-determining regions of the murine ectropic and amphotropic retrovirus envelope proteins. Journal of Virology, 67(8), 4712–4721.

Mortuza, G. B., Dodding, M. P., Goldstone, D. C., Haire, L. F., Stoye, J. P., & Taylor, I. A. (2008). Structure of B-MLV capsid amino-terminal domain reveals key features of viral tropism, gag assembly and core formation. Journal of Molecular Biology, 376(5), 1493–1508.

Moelling K., Broecker F (2015) The reverse transcriptase-RNase H: from viruses to antiviral defense. Ann N Y Acad Sci. 2015 Apr;1341:126-35.

Negroni, M. & Buc, H. (2000) Copy-choice recombination by reverse transcriptases: reshuffling of genetic markers mediated by RNA chaperones. Proc. Natl Acad. Sci. USA 97, 6385–6390.

Nevins , J.R. (2007) Cell transformation by Viruses. In D.N. Knipe and P. M. Howledy, ed Field's Virology (Philadephia: Lippincott-Raven Press), pp.209-248

Nicklas, W., Kraft, V., & Meyer, B. (1993). Contamination of transplantable tumors, cell lines, and monoclonal antibodies with rodent viruses. Laboratory Animal Science, 43(4), 296–300.

Nussbaum, O., Roop, A., & Anderson, W. F. (1993). Sequences determining the pH dependence of viral entry are distinct from the host range-determining region of the murine ecotropic and amphotropic retrovirus envelope proteins. Journal of Virology, 67(12), 7402–7405.

Oakes, B., Tai, A. K., Cingöz, O., Henefield, M. H., Levine, S., Coffin, J. M., & Huber, B. T. (2010). Contamination of human DNA samples with mouse DNA can lead to false detection of XMRV-like sequences. Retrovirology, 7, 109.

Ogura, H. (1976). XC cell fusion by murine leukaemia viruses: fusion from without. Medical Microbiology and Immunology, 162(3-4), 175–181.

Ohkura, S., Goldstone, D. C., Yap, M. W., Holden-Dye, K., Taylor, I. A., & Stoye, J. P. (2011). Novel escape mutants suggest an extensive TRIM5α binding site spanning the entire outer surface of the murine leukaemia virus capsid protein. PLoS Pathogens, 7(3), e1002011.

Omoto, M., Miyashita, H., Shimmura, S., Higa, K., Kawakita, T., Yoshida, S., … Tsubota, K. (2009). The use of human mesenchymal stem cell-derived feeder cells for the cultivation of transplantable epithelial sheets. Investigative Ophthalmology & Visual Science, 50(5), 2109–2115.

Onlamoon, N., Das Gupta, J., Sharma, P., Rogers, K., Suppiah, S., Rhea, J., … Villinger, F. (2011). Infection, viral dissemination, and antibody responses of rhesus macaques exposed to the human gammaretrovirus XMRV. Journal of Virology, 85(9), 4547–4557.

Ovalle Bracho, C., Porras de Quintana, L., Muvdi Arenas, S., & Rios Parra, M. (2007). Polymerase chain reaction with two molecular targets in mucosal leishmaniasis' diagnosis: a validation study. Memórias Do Instituto Oswaldo Cruz, 102(5), 549–554.

Pear, W. S., Nolan, G. P., Scott, M. L., & Baltimore, D. (1993). Production of high-titer helper-free retroviruses by transient transfection. Proceedings of the National Academy of Sciences of the United States of America, 90(18), 8392–8396.

Parthasarathi, S., A, V.-E., Y, R., Bd, P., & Jp, D. (1995). Genetic rearrangements occurring during a single cycle of murine leukaemia virus vector replication: characterization and implications. Journal of Virology, 69(12), 7991–8000.

Patch CT, Chattopadhyay SK, Hauser J, Levine AS. (1981) Regulation of viral transcription in cells infected with iododeoxyuridine-substituted simian virus 40 as a model for the activation by iododeoxyuridine of latent viral genomes. CancerRes. 41(6):2421-7.

Pedersen, L., Behnisch, W., Schmidt, J., Luz, A., Pedersen, F. S., Erfle, V., & Strauss, P. G. (1992). Molecular cloning of osteoma-inducing replication-competent murine leukaemia viruses from the RFB osteoma virus stock. Journal of Virology, 66(10), 6186–6190.

Peebles, P. T., Bassin, R. H., Haapala, D. K., Phillips, L. A., Nomura, S., & Fischinger, P. J. (1971). Rescue of murine sarcoma virus from a sarcoma-positive leukaemia-negative cell line: requirement for replicating leukaemia virus. Journal of Virology, 8(5), 690–694.

Peliska, J. A. & Benkovic, S. J. (1992) Mechanism of DNA strand transfer reactions catalysed by HIV-1 reverse transcriptase. Science 258, 1112–1118.

Pellegrini, G., Traverso, C. E., Franzi, A. T., Zingirian, M., Cancedda, R., & De Luca, M. (1997). Long-term restoration of damaged corneal surfaces with autologous cultivated corneal epithelium. The Lancet, 349(9057), 990–993.

Pera, M. F., Reubinoff, B., & Trounson, A. (2000). Human embryonic stem cells. Journal of Cell Science, 113 (Pt 1), 5–10.

Pizzato, M., Erlwein, O., Bonsall, D., Kaye, S., Muir, D., & McClure, M. O. (2009). A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for

the quantitation of retroviruses in cell culture supernatants. Journal of Virological Methods, 156(1-2), 1–7.

Pomerat, C. M., Kent, S. P., & Logie, L. C. (1957). Irradiation of cells in tissue culture. I. Giant cell induction in strain cultures versus elements from primary explants. Zeitschrift Für Zellforschung Und Mikroskopische Anatomie (Vienna, Austria: 1948), 47(2), 158–174.

Prochazka M, et al. (1991), Emv-30: a novel ecotropic MuLV present in NOD/Lt and NON/Lt strains. Mouse Genome. 89:258

Provitera, P., Goff, A., Harenberg, A., Bouamr, F., Carter, C., & Scarlata, S. (2001). Role of the major homology region in assembly of HIV-1 Gag. Biochemistry, 40(18), 5565–5572.

Quinlan, A. R., Boland, M. J., Leibowitz, M. L., Shumilina, S., Pehrson, S. M., Baldwin, K. K., & Hall, I. M. (2011). Genome Sequencing of Mouse Induced Pluripotent Stem Cells Reveals Retroelement Stability and Infrequent DNA Rearrangement during Reprogramming. Cell Stem Cell, 9(4), 366–373.

Rama, P., Bonini, S., Lambiase, A., Golisano, O., Paterna, P., De Luca, M., & Pellegrini, G. (2001). Autologous fibrin-cultured limbal stem cells permanently restore the corneal surface of patients with total limbal stem cell deficiency. Transplantation, 72(9), 1478–1485.

Ranjith, G. (2005). Epidemiology of chronic fatigue syndrome. Occupational Medicine (Oxford, England), 55(1), 13–19.

Rasheed, S., Gardner, M. B., & Chan, E. (1976). Amphotropic host range of naturally occuring wild mouse leukaemia viruses. Journal of Virology, 19(1), 13–18.

Ravichandran, V., Major, E. O., Ibe, C., Monaco, M. C., Girisetty, M. K. H., & Hewlett, I. K. (2011). Susceptibility of human primary neuronal cells to Xenotropic Murine Leukemia Virus-related (XMRV) virus infection. Virology Journal, 8(1), 443.

Reilly, C. A., & Finkel, M. P. (1975). In vivo interference of virus-induced osteosarcomas by a benign bone tumor virus. Bibliotheca Haematologica, (43), 441–444.

Reuss FU, Frankel WN, Moriwaki K, Shiroishi T, Coffin JM. (1996) Genetics of intracisternal -A- particle-related envelope-encoding proviral elements in mice. J Virol. Sep;70(9):6450-4.

Robert, P.-Y. R., Adenis, J.-P., Denis, F., & Ranger-Rogez, S. (2005). Transmission of viruses through corneal transplantation. Clinical Laboratory, 51(7-8), 419–423.

Robinson, M. J., Erlwein, O., & McClure, M. O. (2011). Xenotropic murine leukaemia virus-related virus (XMRV) does not cause chronic fatigue. Trends in Microbiology.

Robinson, M. J., Tuke, P. W., Erlwein, O., Tettmar, K. I., Kaye, S., Naresh, K. N., … McClure, M. O. (2011b). No Evidence of XMRV or MuLV Sequences in Prostate Cancer, Diffuse Large B-Cell Lymphoma, or the UK Blood Donor Population. Advances in Virology, 2011, 1–6.

Roe T., Reynolds T.C., Yu G., Brown P.O. (1993) Integration of murine leukaemia virus DNA depends on mitosis EMBO J 12(5) pp.2099-2108.

Rosenberg, N., and Jolicoeur, P. (1997) Retroviral pathogenesis. In Retroviruses, Coffin, J.M., Hughes, S.H., and Varmus, H.E. eds Cold Spring Harbor Laboratory Press pp475-585

Rusmevichientong, A., Gupta, J. D., Elias, P. S., Silverman, R. H., & Chow, S. A. (2011). Analysis of Single-Nucleotide Polymorphisms in Patient-Derived Retrovirus Integration Sites Reveals Contamination from Cell Lines Acutely Infected by Xenotropic Murine Leukemia Virus-Related Virus. Journal of Virology, 85(23), 12830–12834.

Sakuma, T., De Ravin, S. S., Tonne, J. M., Thatava, T., Ohmine, S., Takeuchi, Y., … Ikeda, Y. (2010). Characterization of retroviral and lentiviral vectors pseudotyped with xenotropic murine leukaemia virus-related virus envelope glycoprotein. Human Gene Therapy, 21(12), 1665–1673.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., … Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology, 12(1)12:87..

Sato, E., Furuta, R. A., & Miyazawa, T. (2010). An endogenous murine leukaemia viral genome contaminant in a commercial RT-PCR kit is amplified using standard primers for XMRV. Retrovirology, 7, 110.

Schlaberg, R., Choe, D. J., Brown, K. R., Thaker, H. M., & Singh, I. R. (2009). XMRV is present in malignant prostatic epithelium and is associated with prostate cancer, especially high-grade tumors. Proceedings of the National Academy of Sciences of the United States of America, 106(38), 16351–16356.

Schloss, P. D., Gevers, D., & Westcott, S. L. (2011). Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. PLoS ONE, 6(12), e27310.

Schuster SC. (2008). Next-generation sequencing transforms today's biology. Nat Methods. 2008, 5 (1): 16-18. Nat Methods. 5 (1): 16-18.

Schwab, I. R., Reyes, M., & Isseroff, R. R. (2000). Successful transplantation of bioengineered tissue replacements in patients with ocular surface disease. Cornea, 19(4), 421–426.

Seiki, M., Hattori, S., Hirayama, Y., & Yoshida, M. (1983). Human adult T-cell leukaemia virus: complete nucleotide sequence of the provirus genome integrated in leukaemia cell DNA. Proceedings of the National Academy of Sciences of the United States of America, 80(12), 3618–3622.

Sfanos, K. S., Aloia, A. L., De Marzo, A. M., & Rein, A. (2012). XMRV and prostate cancer-a “final” perspective. Nature Reviews. Urology. Jan 10;9(2):111-8

Sharma D, Priyadarshini P, Vrati S. (2015) Unraveling the web of viroinformatics: computational tools and databases in virus research. J Virol. 89(3):1489-501.

Sharp, P. M., Bailes, E., Robertson, D. L., Gao, F. & Hahn, B. H. (1999) Origins and evolution of AIDS viruses. Biol. Bull. 196, 338–342

Sharp PM, Hahn BH (2011) Origins of HIV and the AIDS epidemic. Cold Spring Harb Perspect Med. Sep;1(1)

Shin, C. H., Bateman, L., Schlaberg, R., Bunker, A. M., Leonard, C. J., Hughen, R. W., … Singh, I. R. (2011). Absence of XMRV Retrovirus and Other Murine Leukemia Virus-Related Viruses in Patients with Chronic Fatigue Syndrome. Journal of Virology, 85(14), 7195–7202.

Shortt, A. J., Secker, G. A., Rajan, M. S., Meligonis, G., Dart, J. K., Tuft, S. J., & Daniels, J. T. (2008). Ex Vivo Expansion and Transplantation of Limbal Epithelial Stem Cells. Ophthalmology, 115(11), 1989–1997.

Simmons, G., Glynn, S. A., Komaroff, A. L., Mikovits, J. A., Tobler, L. H., Hackett, J., … for the Blood XMRV Scientific Research Working Group (SRWG). (2011). Failure to Confirm XMRV/MLVs in the Blood of Patients with Chronic Fatigue Syndrome: A Multi-Laboratory Study. Science.Mar 17;8:20.

Smith, R. A. (2010). Contamination of clinical specimens with MLV-encoding nucleic acids: implications for XMRV and other candidate human retroviruses. Retrovirology, 7, 112.

Snodgrass, R. G., Collier, A. C., Coon, A. E., & Pritsos, C. A. (2010). Mitomycin C inhibits ribosomal RNA: a novel cytotoxic mechanism for bioreductive drugs. The Journal of Biological Chemistry, 285(25), 19068–19075.

Sørensen, K. D., Kunder, S., Quintanilla-Martinez, L., Sørensen, J., Schmidt, J., & Pedersen, F. S. (2007). Enhancer mutations of Akv murine leukaemia virus inhibit the induction of mature B-cell lymphomas and shift disease specificity towards the more differentiated plasma cell stage. Virology, 362(1), 179–191.

Stephenson, J. R., Crow, J. D., & Aaronson, S. A. (1974). Differential activation of biologically distinguishable endogenous mouse type C RNA viruses: interaction with host cell regulatory factors. Virology, 61(2), 411–419.

Stevens, A., Bock, M., Ellis, S., LeTissier, P., Bishop, K. N., Yap, M. W., … Stoye, J. P. (2004). Retroviral Capsid Determinants of Fv1 NB and NR Tropism. Journal of Virology, 78(18), 9592–9598.

Stieler, K., Schulz, C., Lavanya, M., Aepfelbacher, M., Stocking, C., & Fischer, N. (2010). Host range and cellular tropism of the human exogenous gammaretrovirus XMRV. Virology, 399(1), 23–30.

Stocking C., and Kozak, C.A. (2008) Murine endogenous retroviruses. Cell. Mol. Life Sci. 65;3383-3398

Stoye, J. P., & Coffin, J. M. (1987). The four classes of endogenous murine leukaemia virus: structural relationships and potential for recombination. Journal of Virology, 61(9), 2659–2669.

Stoye, J. P., Moroni, C., & Coffin, J. M. (1991). Virological events leading to spontaneous AKR thymomas. Journal of Virology, 65(3), 1273–1285.

Stoye, J. P. (2006). Koala retrovirus: a genome invasion in real time. Genome Biology, 7(11), 241.

Stoye, J. P. (2012). Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nature Reviews. Microbiology, 10(6), 395–406.

Strong, M. J., Xu, G., Morici, L., Splinter Bon-Durant, S., Baddoo, M., Lin, Z., … Flemington, E. K. (2014). Microbial Contamination in Next Generation Sequencing: Implications for Sequence-Based Analysis of Clinical Samples. PLoS Pathogens, 10(11), e1004437.

Sung, W.-K. et al. (2012) Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nat. Genet. 44, 765–769.

Sutcliffe, J.G., Shinnick, T.M., Verma, I.M., and Lerner R.A. (1980) Nucleotide sequence of Moloney Leukaemia virus : 3' end reveals details of replication, analogy to bacterial transposons and an unexpected gene. Proc. Natl. Acad. Sci. USA 77, 3302-3306

Switzer, W. M., Jia, H., Hohn, O., Zheng, H., Tang, S., Shankar, A., … Heneine, W. (2010). Absence of evidence of xenotropic murine leukaemia virus-related virus infection in persons with chronic fatigue syndrome and healthy controls in the United States. Retrovirology, 7, 57.

Switzer, W. M., Zheng, H., Simmons, G., Zhou, Y., Tang, S., Shankar, A., … Heneine, W. (2011). No evidence of murine leukaemia virus-related viruses in live attenuated human vaccines. PloS One, 6(12), e29223.

Takeuchi, T., Wang, L., Mori, S., Nakagawa, K., Yoshikura, H., & Kanda, T. (2008a). Characterization of mouse 3T3-swiss albino cells available in Japan: necessity of quality control when used as feeders. Japanese Journal of Infectious Diseases, 61(1), 9–12.

Takeuchi, Y., McClure, M. O., & Pizzato, M. (2008b). Identification of gammaretroviruses constitutively released from cell lines used for human immunodeficiency virus research. Journal of Virology, 82(24), 12585–12588.

Teich, N. M., Weiss, R. A., Martin, G. R., & Lowy, D. R. (1977). Virus infection of murine teratocarcinoma stem cell lines. Cell, 12(4), 973–982.

Teofili, L., Iachininoto, M. G., Capodimonti, S., Ucciferri, C., Nuzzolo, E. R., Martini, M., … Larocca, L. M. (2010). Endothelial progenitor cell trafficking in human immunodeficiency virus-infected persons. AIDS, 24(16), 2443–2450.

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., & Jones, J. M. (1998). Embryonic Stem Cell Lines Derived from Human Blastocysts. Science, 282(5391), 1145–1147.

Todaro, G. J., De Larco, J. E., & Cohen, S. (1976). Transformation by murine and feline sarcoma viruses specifically blocks binding of epidermal growth factor to cells. Nature, 264(5581), 26–31.

Todaro, G. J., & Green, H. (1963). Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. The Journal of Cell Biology, 17, 299–313.

Topping, R., Demoitie, M. A., Shin, N. H., & Telesnitsky, A. (1998). Cis-acting elements required for strong stop acceptor template selection during Moloney murine leukaemia virus reverse transcription. Journal of Molecular Biology, 281(1), 1–15.

Triviai, I., Ziegler, M., Bergholz, U., Oler, A. J., Stübig, T., Prassolov, V., … Stocking, C. (2014a). Endogenous retrovirus induces leukaemia in a xenograft mouse model for primary myelofibrosis. Proceedings of the National Academy of Sciences of the United States of America, 111(23), 8595–8600.

Triviai, I., Ziegler, M., Bergholz, U., Oler, A. J., Stübig, T., Prassolov, V., … Stocking, C. (2014b). Endogenous retrovirus induces leukaemia in a xenograft mouse model for primary myelofibrosis. Proceedings of the National Academy of Sciences of the United States of America, 111(23), 8595–8600.

Trono, D. (1995). HIV accessory proteins: leading roles for the supporting cast. Cell, 82(2), 189–192.

Tuke, P. W., Tettmar, K. I., Tamuri, A., Stoye, J. P., & Tedder, R. S. (2011). PCR master mixes harbour murine DNA sequences. Caveat emptor! PloS One, 6(5), e19953.

Uren, A. G., Mikkers, H., Kool, J., van der Weyden, L., Lund, A. H., Wilson, C. H., Adams, D. J. (2009). A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. Nature Protocols, 4(5), 789–798.

Urisman, A., Molinaro, R. J., Fischer, N., Plummer, S. J., Casey, G., Klein, E. A., DeRisi, J. L. (2006). Identification of a novel Gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. PLoS Pathogens, 2(3), e25.

van der Kuyl, A. C., Dekker, J. T., & Goudsmit, J. (1999). Discovery of a new endogenous type C retrovirus (FcEV) in cats: evidence for RD-114 being an FcEV(Gag-Pol)/baboon endogenous virus BaEV(Env) recombinant. Journal of Virology, 73(10), 7994–8002.

Van der Rauwelaert, E., Maisin, J. R., & Merregaert, J. (1988). Provirus integration and myc amplification in 90Sr induced osteosarcomas of CF1 mice. Oncogene, 2(3), 215–222.

van Kuppeveld, F. J. M., de Jong, A. S., Lanke, K. H., Verhaegh, G. W., Melchers, W. J. G., Swanink, C. M. A., … van der Meer, J. W. M. (2010). Prevalence of xenotropic murine leukaemia virus-related virus in patients with chronic fatigue syndrome in the Netherlands: retrospective analysis of samples from an established cohort. BMJ (Clinical Research Ed.), 340, c1018.

van Kuppeveld, F. J., & van der Meer, J. W. (2012). XMRV and CFS—the sad end of a story. The Lancet. Feb 4;379(9814):e27-8.

Vaughan AE, Mendoza R, Aranda R, Battini JL, Miller AD. (2012) Xpr1 is an atypical G-protein-coupled receptor that mediates xenotropic and polytropic murine retrovirus neurotoxicity. J Virol. Feb; 86(3):1661-9.

Vogt M. (1979) Properties of "mink cell focus-inducing" (MCF) virus isolated from spontaneous lymphoma lines of BALB/c mice carrying Moloney leukaemia virus as an endogenous virus. Virology. 93(1):226-36

Varela-Echavarría, A., Garvey, N., Preston, B. D., & Dougherty, J. P. (1992). Comparison of Moloney murine leukaemia virus mutation rate with the fidelity of its reverse transcriptase in vitro. The Journal of Biological Chemistry, 267(34), 24681–24688.

Voelkel, C., Galla, M., Maetzig, T., Warlich, E., Kuehle, J., Zychlinski, D., … Baum, C. (2010). Protein transduction from retroviral Gag precursors. Proceedings of the National Academy of Sciences, 107(17), 7805–7810.

Wang, Q., Jia, P. & Zhao, Z. (2013) VirusFinder: software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. PLoS One 8, e64465.

Weiss, R. A. (2010). A cautionary tale of virus and disease. BMC Biology, 8, 124.

Wensel, D. L., Li, W., & Cunningham, J. M. (2003). A Virus-Virus Interaction Circumvents the Virus Receptor Requirement for Infection by Pathogenic Retroviruses. Journal of Virology, 77(6), 3460–3469.

Wilk, T., de Haas, F., Wagner, A., Rutten, T., Fuller, S., Flügel, R. M., & Löchelt, M. (2000). The intact retroviral Env glycoprotein of human foamy virus is a trimer. Journal of Virology, 74(6), 2885–2887.

Willis M.C., Hicke,B.J., Uhlenbeck,O.C., Cech,T.R. and Koch,T.H. (1993) Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. Science, 262, 1255–1257.

Wolff, D., & Gerritzen, A. (2011). Presence of murine leukaemia virus (MLV)-related virus gene sequences in a commercial RT-PCR reagent. Clinical Laboratory, 57(7-8), 631–634.

Wolf, G., Yang, P., Füchtbauer, A. C., Füchtbauer, E.-M., Silva, A. M., Park, C., … Macfarlan, T. S. (2015). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. Genes & Development, 29(5), 538–554.

Wu, X., Li, Y., Crise, B., & Burgess, S. M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science (New York, N.Y.), 300(5626), 1749–1751.

Wu, X., Li, Y., Crise, B., Burgess, S. M., & Munroe, D. J. (2005). Weak Palindromic Consensus Sequences Are a Common Feature Found at the Integration Target Sites of Many Retroviruses. Journal of Virology, 79(8), 5211–5214.

Yamashita M., Emerman M. (2006) Retroviral infection of non-dividing cells: old and new perspectives. Virology 5;344(1):88-93.

Yang K, Lee J, Cho SW. (2012) Engineering biomaterials for feeder-free maintenance of human pluripotent stem cells. Int J Stem Cells. May;5(1):1-5.

Young, J.A. (2001) Virus entry and uncoating. In Field's Virology, Knipe, D.M., ed (Philadephiea: Lippincott Williams &Wilkins), pp. 87-103

Young, G., Kassiotis, G., & Stoye, J. (2012a). Emv2, the only endogenous ecotropic murine leukaemia virus of C57BL/6J mice. Retrovirology, 9(1), 23.

Young, G. R., Eksmond, U., Salcedo, R., Alexopoulou, L., Stoye, J. P., & Kassiotis, G. (2012b). Resurrection of endogenous retroviruses in antibody-deficient mice. Nature, 491(7426), 774–778.

Yu, S. S., Han, E., Hong, Y., Lee, J.-T., Kim, S., & Kim, S. (2003). Construction of a retroviral vector production system with the minimum possibility of a homologous recombination. Gene Ther, 10(8), 706–711.

Yu SC, Xu YY, Li Y, Xu B, Sun Q, Li F, Zhang XG. (2015) Construction of tissue engineered skin with human amniotic mesenchymal stem cells and human amniotic epithelial cells. Eur Rev Med Pharmacol Sci. 2015 Dec;19(23):4627-35.

Zanetti, G., Briggs, J. A. G., Grünewald, K., Sattentau, Q. J., & Fuller, S. D. (2006). Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. PLoS Pathogens, 2(8), e83.

Závada J (1982) The pseudotypic paradox J Gen Virol. Nov; 63:15-24.

Zhang, J., & Temin, H. M. (1993). Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. Science (New York, N.Y.), 259(5092), 234–238.

Zhang, Y.-A., Maitra, A., Hsieh, J.-T., Rudin, C. M., Peacock, C., Karikari, C., … Gazdar, A. F. (2011). Frequent detection of infectious xenotropic murine leukaemia virus (XMLV) in human cultures established from mouse xenografts. Cancer Biology & Therapy, 12(7).

APPENDIX

	Name	sequence 5'-3'	bp	start	end	gene	orient.
MLV target (PCRs)							
	GagOuterF	ggatcaagccccacatacag	20	3294	3313	*gag*	FW
	GagInnerF	agaagtcaacaagcgggtgg	20	3423	3443	*gag*	FW
	GagOuterR	catcaaacagggtgggactg	20	3658	3677	*gag*	REV
	GagInnerR	ggtggagtctcaggcagaaa	20	3541	3560	*gag*	REV
	GagProbe	caccgtgcccaacccttacaacctct	26	2964	2989	*gag*	probe
	XMRVgag F	accgtttgtctctcctaaac	20	906	925	*gag*	FW
	XMRVgag R	agggtaaagggcagatcg	18	1475	1492	*gag*	REV
	XMRVgag probe	ccgacagctcccgtcctcccg	21	937	957	*gag*	probe
	env FOR	tcctgctgaccacccccacc	20	6146	6166	*env*	FW
	env REV	gatgaggatggcttccag	18	7030	7012	*env*	REV
	mRNAENV-F	gtctttcatttgggggctcg	20	601	621	U5-LTR	FW
	mRNAENV-R	tcctttccagcgaggttcta	20	6315	6335	*env*	REV
	mRNAENV-probe	ccagtgataccacacccctt	20	6249	6269	*env*	probe
	AKV_BC	ggactaagggccaaacaggat	21	8043	8064		FW
	KiMSVF	agctagtggcgtaggcaaga	20	3062	3083	*p15*	FW
	KiMSVR	aactcctgagcctgtttcgt	20	3150	3171	*p15*	REV
Oligonucleotides splikerette							
	long adaptor	cgaagagtaaccgttgctaggagagaccgtggct gaatgagactggtgtcgacactagtgg	61	-	-	-	-
	short adaptor	gatcccactagtgtcgacaccagtctctaatttttttttt tcaaaaaaa	48	-	-	-	-
	splink1	cgaagagtaaccgttgctaggagagacc	28	-	-	-	-
	splink2	gtggctgaatgagactggtgtcgac	25	-	-	-	-
	CF1 outer	gagaggctggaaagtaccgg	20	115	134	U3_LTR	REV
	CF1 inner	gccattttgcaaggcatgga	20	47	66	U3_LTR	REV
	XMRV outer	cccctgtcccagtgttattcagcctta	27	7840	7866	U3_LTR	REV
	XMRV inner	tgtaacttttgagaactcagctctggt	27	7794	7820	U3_LTR	REV
	MoMLV outer	gcgttacttaagctagcttgccaaacctac	30	7569	7598	U3_LTR	REV
	MoMLV inner	ccaaacctacaggtggggtctttc	24	7555	7578	U3_LTR	REV
Internal control and reference gene targets							
	MS2 For	gtacgggcgaccccacgatgac	22	-			
	MS2 REV	tggcactacccctctccgtattcacg	26				
	MS2 probe	cacatcgatagatcaaggtgcctacaagc	29				

	m_HPRT1F	gcaaactttgctttccctgg	20				
	m_HPRT1R	acttcgagaggtccttttcacc	22				
	mHPRT1_probe	cgccccaaaatggttaaggttgcaag	27				

Appendix 2 Methods (PERT) optimisation

False positive results arise in PERT due to contaminating enzyme in sample supernatant. Having controlled for inadvertent environmental contamination with the disposal of tips, one-off use of buffer aliquots, extra sterilisation measures and minimisation of handling (following the contamination prevention measures of Kwok and Higuchi ,1989); the following may also be sources of false RT activity in cell culture:

a) RT-like activity of DNA dependent Taq polymerase
b) Cellular polymerase noise inhibition
c) Residual RNA in the PCR reaction

RT-like activity of DNA dependent Taq polymerases

There is a widely documented but unproven knowledge that the DNA dependent Taq-polymerase used in Taq assays may amplify residual RNA from cDNA transferral. Indeed, direct testing of the Taq polymerase from the second step of the PERT assay in the cDNA synthesis step, I demonstrated that the DNA dependent enzyme was indeed exerting a significant RNA- dependent action in the presence of residual RNA. The directly amplifying action of the tested Taq polymerase (GoTaq Promega M5005) on MS2 RNA is shown in Appendix Figure 2.1B . In the figure, it is shown that 1 IU of Taq Polymerase can amplify 48ng of MS2 RNA when directly used in the PERT assay as a putative RT sample, at a level comparable to an Mo-MLV RT quantity of 10^{-6} IU; while this activity can be significantly suppressed by activated DNA. Two further dilutions of the Taq polymerase enzyme tested, demonstrated that the RT-like activity is dependent on the quantity of the polymerase enzyme proportionately. Other known cases of this RT-like activity arise in the literature (Appendix Table 2.1) that however are not quantified.

Cellular polymerase noise

Cellular polymerases such as polymerase A (DNA dependent), telomerases and retrotransposons (Maudru et al, 1997) also exert RT-like action on the template RNA. This has traditionally been addressed by authors with the use of cellular polymerase inhibitors such as activated DNA from calf thymus, heparin or phonoacetic acid (for viral polymerases) or shrimp DNase (Amersham Biosciences, cat. no. US78314). Others have evaluated lowering the assay pH.

To mimic the action of cellular polymerase noise, I obtained a stock of polymerase 1 (Promega, M 205A), which was subsequently used in inhibition studies as the most proximate polymerase to the polymerases in cell extracts.

I evaluated calf thymus activated DNA on the spiking polymerase enzyme (Polymerase 1, M 205A Promega) and on cell culture supernatants for its ability to inhibit cellular polymerase noise. The activated DNA inhibition of polymerase activity is sought to occur with at least 1:104 minimum ratio of RNA template/activated DNA (Lovatt 1999). McCormick et al have used 1:300 (20ng/6µg) and observe a one million-fold suppression in polymerase activity. In both these published assays however, the effect of activated DNA suppression on the standard and sample true RT enzymatic activity is not evaluated.

I titrated and optimised the addition of activated DNA in variable ratios to the template RNA present in the cDNA synthesis step and found that there is complete abolishment of the signal obtained from the same amount of spiking polymerase 1 at ratios above 1/80 RNA/aDNA confirming the previous literature. However, although ratios of 1/40 to 1/130 of activated DNA appear to not affect the amplification of native Mo-MLV RT polymerase in the original dilutions of 10^{-4} and 10^{-6} IU of Mo-MLV standard/ reaction, they consistently abolished the lower Mo-MLV RT concentrations (10^{-8} and lower) thus affecting the assay sensitivity. At ratios of 1/200 to 1/500, the amplification of the RT standard dilution of 10^{-4} was also shifted significantly (3log10 reduction to complete abolishment). Transferred activated DNA was also found to

affect the fluorescence signal quality (ΔRn) in all the conditions tested. This finding partially contradicts the outcomes from McCormick et al, who reported that a ratio of 1/300 activated DNA did not significantly affect the standard detection of MoMLV-RT in their settings.

The addition of activated DNA was therefore, not used in our experiments to avoid inhibition of low level true retroviral activity.

RNA input quantity

There is significant variation in the literature for the amount of RNA that is used as an optimal template for reverse transcriptase detection and quantitation in PERT. For example, Ma et al 2009 uses 240ng of RNA template in each cDNA synthesis well (25 µl) in a protocol that is directed for cell culture. Fan et al, 2006 uses an even larger amount of 300ng/reaction. In Brorson et al, 2008 for the detection of RTase in monocolonal antibody cell culture, the authors use 6ng of MS2 RNA per well in a 25µl cDNA synthesis mix.The last trend i.e using low quantities of RNA, is also observed in similar PERT assays that use BMV RNA, with quantities as low as 120ng (Pizzato, 2009), 20ng (Mc Cormick 2008) and 12ng (Lovatt 1999) per reaction well (20µl, 20µl and 15µl respectively).

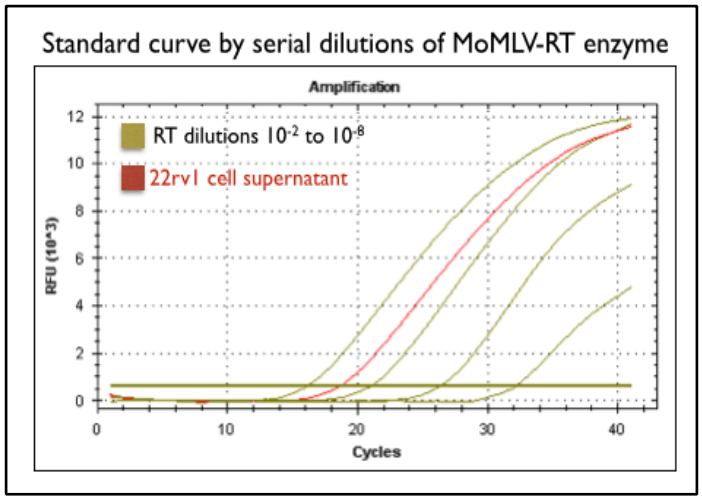
The RNA added as a template is paramount for the action of the putative RT and the commercial RT used for the creation of the standard curve. However, using too much and most of the RNA will be left unused, which means that the excess will be transferred inadvertently to the PCR mix at the cDNA transferral stage, creating noise which is considered to be proportionate to that remnant quantity (Lovatt et al, 1999). Indeed, Lovatt et al, managed to completely abolish the signal arising from cell-associated polymerases in cell lysates by reducing the BMV RNA level from 60ng to 15ng and below per reaction well.

We therefore, titrated the input RNA template to adjust the assay's performance and in the same time reduce the incidence of false positive RT negative wells. Using this

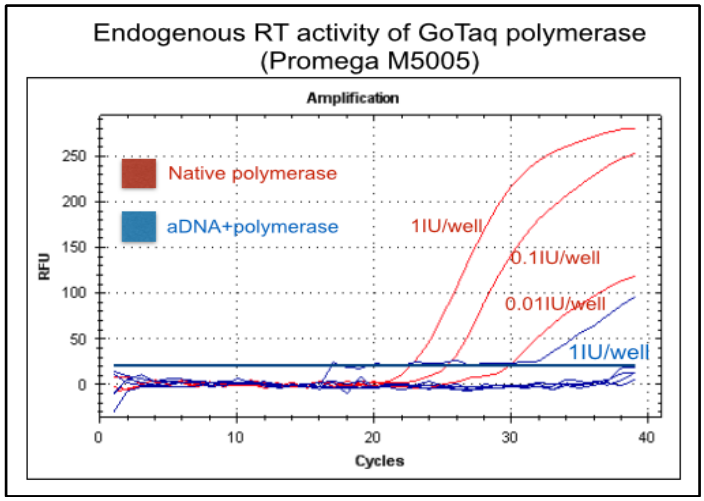
approach, we managed to reduce the amount of starting RNA template in the cDNA synthesis step to 48ng/well in a 25μl reaction. Interestingly as little as 9.6ng of MS2 RNA is able to produce cDNA that is competently amplified, producing identical CT values with higher RNA input, has a satisfactory signal fluorescence and remains unaffected for low RT concentrations. Under the starting 240ng/ well, any 5-fold decrease in the starting template quantity affected the cycle threshold by 2.16 cycles on average (10 fold reduction), while this shift was more distinctive for RT values lower than the usual detection limit (Poisson distribution effect).

Figure Appendix 2.1. Optimisation of the PERT assay: standard curve and background noise strategies

A.



B.



C.

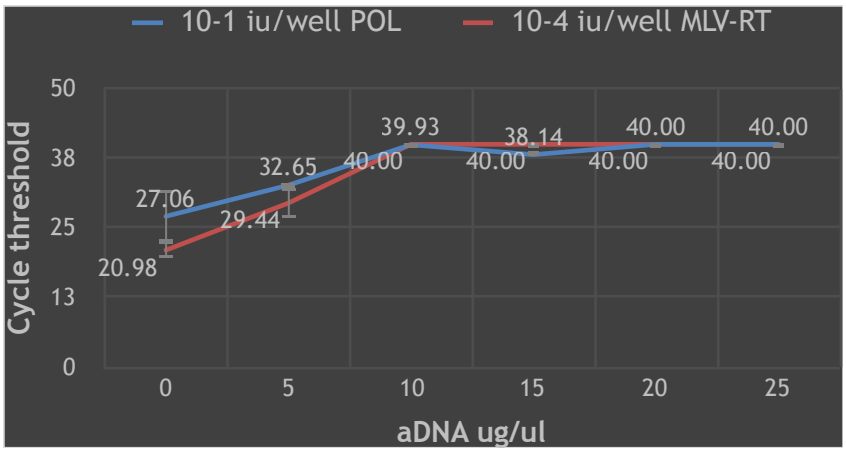


Figure Appendix 2.1 Optimisation of the PERT assay: standard curve and background noise strategies

A) **Standard curve - PERT assay:** Serial dilutions shown in the amplification output of the BioRad RT-PCR CXR, were constructed using a standard purified recombinant MoMLV RT enzyme (200IU/µl) diluted from 10^{-2} to 10^{-8} IU/reaction well (green amplification curves). Lysed supernatant from 22rv1 cells, producing XMRV, was used as a positive cell culture derived RT activity control (red amplification curve). By previous PCR quantification, 10µl of 22rv1 supernatant contain 10^{8} XMRV retroviral particles.

B) **The amplification signal obtained through the direct action of Taq polymerase (GoTaq, Promega M5005) in the PERT assay:** Serial dilutions of DNA dependent Taq polymerase were tested as putatively RT containing samples in the PERT assay (range 1 IU/well to 10^{-6} IU/well). The polymerase was able to amplify MS2 RNA template (48 ng/well) and produce a detectable cDNA in the RT-PCR assay that was quantified (via the quantification curve obtained in figure 2.1A) at a level equivalent to $5x10^{-6}$ IU of RT activity per IU of Taq polymerase. This finding demonstrates how residual RNA in the RT-PCR step may be amplified by Taq polymerase, producing false positives in non-template controls and confounding true retroviral enzymatic activity quantification.

C) **Effect of varied amounts of activated DNA (aDNA) on native polymerase 1 and standard MoMLV RT enzyme:** The action of 0.1IU of polymerase 1 is suppressed from CT 27 to CT 32 (with up to 5 µg/well of aDNA) and to CT 39 (with up to 10 µg/well of aDNA) (red line). A similar shift is observed with a standard dilution of 10^{-4} of MoMLV-RT (shift from CT 20 to CT 29 to CT 40).

Table Appendix 2.1: Commercial polymerase endogenous RT-like activity

Polymerases with endogenous low RT activity		
Commercial name	**company**	**RT activity proof -Reference**
Platinum Taq	Thermo Fisher Scientific	Background signal in 20-30 cycles in negative controls Pizzato 2009
Ampli Taq	Applied Biosystems	Maudru and Peden 1997, Pyra et al,1994
Platinum Taq	Thermo Fisher Scientific	Maudru and Peden 1997
UITMa		Maudru and Peden 1997 (lower with MS2 RNA)
native Pfu	Agilent	Maudru and Peden 1997, (very high with TMV RNA)
Taq, no other info		Jones and Foulkes, 1989
Recomb. Taq	Invitrogen	Fan et al,2006
Recomb.Taq	Qiagen	Fan et al,2006
Recomb.Taq	TaKaRa	Fan et al,2006
Recomb.Taq	Sangon	Fan et al,2006
TTH	Promega	High endogenous RT activity
Polymerases that lack endogenous low RT activity		
Rec.Taq polymerase	-MBI Fermentas -Promega	lacks RT activity: Fan et al,2006 and Maudru and Peden 1997
PFU µltra	Agilent	Arezi et al, 2003
PFU Turbo	Agilent	Arezi et al, 2003
AmpliTaq Gold	Applied Biosystems	Arezi et al, 2003
Easy-A	Agilent	Arezi et al, 2003

Figure Appendix 2.2: PERT assay quantification curve reproducibility

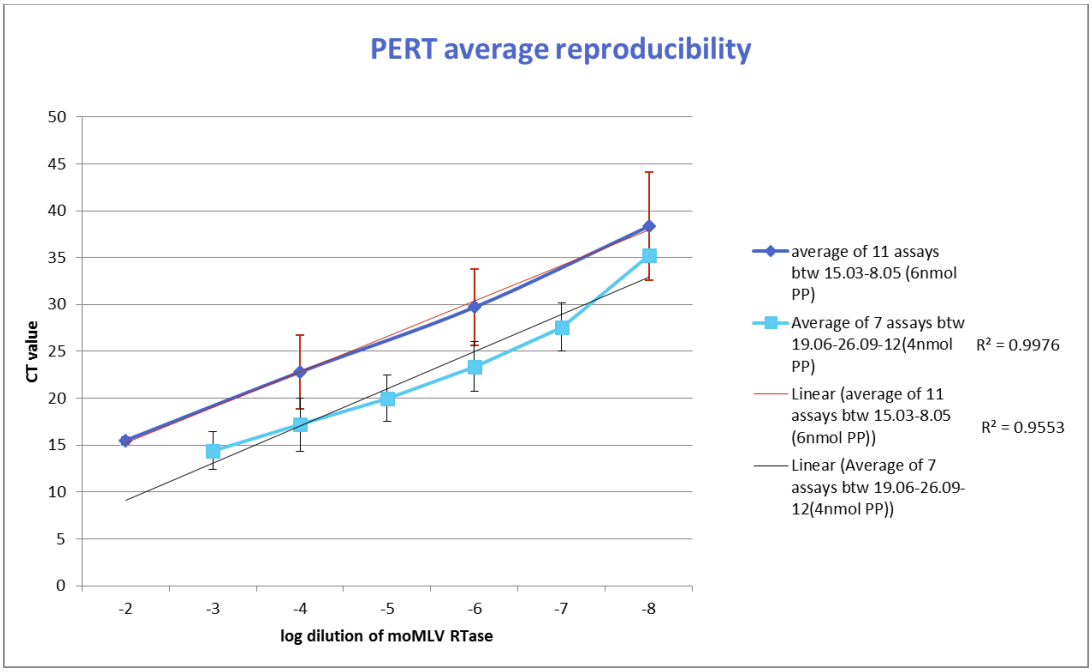


Figure Appendix.2.2 PERT assay quantification curve reproducibility The standard curve results average created with serial dilutions of Mo-MLV RT enzyme for 18 assays clustered in two groups of different primer concentration (6nmol vs 4nmol per µl in PCR reaction), were plotted and the efficiency of the two methods compared. Lower primer concentration results in a higher PCR efficiency (99.76%) versus (95.53%) for higher primer concentration, but not significant difference in the assays limit of detection. However, as in CT values above 30 the background noise of samples is increased with false positives arising on average at CT=31, we valued earlier detection coupled with higher PCR efficiency and optimised our protocol with the lower primer concentration.

Appendix Table 3 Studies evaluating 5-IDU for the induction of MLV

Table 3 Appendix

Inducer	*Cell line*		*Other chemicals*	*Released virus*	*Dose*	*Start Peak End*	*Observations*	*Detection*	Ref
	human	*mouse*							
IdU	HEL-12			C type virus	*30µg/ml*	*Day 25 Day64 n/a*	*Spontaneous at day 80 (control)*	*Anti-RD1I4 p28 and anti-SiSV (TE) electron microscopy and buoyant density.*	[1] 1979
	HBT-3		*+testosterone acetate*	C type virus	*5mg/ml testo acetate for 8 wks IdU 100µg/ml for 3 d*	*Peak at 28 pass*	-	*RNA dependent DNA polymerase activity Not reactive to rat RD-114 or primate sera – not infecting other cell lines*	[2] 1974
		BALB-c and Human rhabdomyo sarcoma		C type and **immature IAP**	*20 µg/ml*	*Wk 2*	*Control 3T3 line latent release IAP*	*XC plaque assay RT activity EM inoculated on to plates of 3T3*	[3] 1975
		CSH/ MCA5 C3H/ST3	-	N-ecotropic Defective MuLV	*n/a*	*11 pass 20 pass n/a*	-	*XC plaque assay*	[4] 1976
		NIH/3T3 SC-1 AKR 2B	*+DNA from AKR virus (-)cells*	Ecotropic-MLV	*40µg/ml AKR(24h) 100µg/ml*	*n/a*	*n/a*	IdUrd readily activates the endogenous ecotropic AKR MuLV from AKR 2B	[5] 1978
		NIH Swiss/ N BALB/c N C58/J	-	C type	*(30 µg/ml) for 24 hr*	*4-5 days 16 days*	-	*Fv-1 and fv-1^n hybrids*	[6] 1974
		NZB/BLN NIH/3T3 and BALB/ 3T12	-	Ki-MLV and R-MLV(NB-tropic)	*20 µg/ml*	*Day 4*	*6 to 8-fold NIH not induced*	*P30 P12*	[7] 1976
		B10.BR/ SgLi	-	B and N tropic MLV	-	-	-	-	[8]
		K-BALB	*+5-AzaC2 µg/ml*	C type and IAP	*30 µg/ml for 24h*	*Day1 n/a*	-	PERT and STFPERT infection of NIH3T3 cells	[9] 2001
		K-BALB	*+glycocorticoids*	Murine type C x-tropic until 5-7 days and then N-tropic	*40 µg/ml for 24h 25 Mg/ml DEAE 1h*	*Day 2 Day3 Day 6*	*Both new and transfected induction*	*Post-transcriptional Cordycepin Interferon RTase and NRK cells (rat)*	[10] 1976
		NIH 3T3 / BALB 3T3	-	Endogenous MLV	*30 µg/ml*	*Day 1-2 -*	*Random batch*		[11] 2008
		BALB/C BALB/K	*+cordycepin +steroids +insulin*	Type C virus	*40µg/ml for 24h Steroids 24h prior*	*- Day3 -*	*1.8-2 fold (cortis) 25- fold (dexameth)*	*The stimulation in nonproducing cells ranges from 2-fold to 25-fold (dexamethazone)*	[12] 1973
		BALB/c C58	*Steroids Cycloheximide Mitomycin C*	R-MLV		*24h intervals*	*2.5 fold increase first 48h by hydrocortisone*		[13] 1975
		BALB-K	-	After: Mitomycin C 25 ug/ml for 60 min.	*30µg/ml for 20h at 37o*	*- Day3*	*Plateau at 10µg/ml*	*Spontaneous also (BALB-2) infectious center assay on NRK*	[14] 1974
		BALB-c	*+cycloheximide +HU +Mitomycin C +cytosine arabinoside*	Xenotropic type C	*10 µg/ml cycloheximide for 18 h*	*24 -29h cyclohex 8 to 13h IdU*		*(3.0%) with IdU*	[15] 1975
		MLg	*+chemical carcinogens +microsomal enzymes +NADPH2*	Endogenous type C	-	-	*10fold increase*	-	[16] 1976
		MLg	-	N-tropic,	25µg/ml	*24h*	-	*76th passage cells released the virus 'spontaneously'*	[17] 1975
		Neuroblastoma clone N4,	*dexomethosazone*	Type C		*- Day 3 Day5*	*C-particle antigen rose 20- to 30-fold per milligram of cell protein*	-	[18] 1976
		AKR	*Or BrDU*	MLV	*20-100 µg*	*Day 3 - Day14*	*0.1 to 0.5% of cells exposed*	-	[19] 1971

Inducer	Cell line		Other chemicals	Released virus	Dose	Start Peak End	Observations	Detection	Ref
	human	mouse							
		AKR	*Compounds* +irradiation +BrDU*	MLV	*several*	*Day 7 Day14*	*Irradiation enhanced virus activation*		[20] 1973
		AKR BALB/c NIH/3T3	*BP, benzo (a) pyrene ; BPD, 7.8-diol-benzo (a) pyrene; BPDE, 7.8-diol-9, 10-epoxy-benzo(a)pyrene; AZA, 5-azacytidine*	Endogenous RVs	*IdUrd varied from 20 to 50 µg/ml [3H]BP was added at a final concentration of 100 mM*	*Day 2 and day 3 And day 4(AZA) Declined by day6*	*5-azacytidine induced >90% of AKR cells at a concentration of 2 to 4 µg/ml*	*Mainly comparison of 5-IDU and others*	[21] 1982
		AKR	Cambamoyloxyurea Formamidoxime Hydroxyurea	Endogenous murine virus	*HU 2mM IdURD 50µg/ml*	*Day 9 Passage 6 -*	*0.01-0.03% induction*	*SC1 cells lack the fv-1 restriction gene*	[22] 1978
		C3H/ 1OT1/2. C3H/ST3, C3H/ MCA5, and AKR BALB/ 3T3, SC-1 rabbit cornea SIRC	-	N-tropic in transformed and replication defective in C3H/10T1/2	*IdU 20µg/ml for 24h*			*XC virus plaque assay Xenotropic MuLV was not detected*	[23] 1976
		C57BL/6N, NIH/N, and BALB/cN			20 µg/ml 24h	Peak 3-4 days			[24] 1972
		AKR	0.5 to I µM hydrocortisone	MLV	IdU 100 µg/ml for 24 hr		hydrocortisone modulates the extent of expression	serum starved for 8 hr and then treated with IdUrd for 24 hr, with or without hydrocortisone.	[25] 1974
		AKR-2B (a highly ldUrd-inducible subclone of 32C) (35); BALB/c 3T3 and NIH 3T3	-	Endogenous MLV	(20 µg/ml,24 hr)			Induction of the transcription of the AKR genome (Radiolabelled nucleotides)	[26] 1979
		AKR C3H/He and BALB/c		Endogenous MLV	50 ug of IdUrd per ml for 18 h,	chronic productive infection	viral p30 antigen production over a period of 42 days	RTase and XC plaque	[27] 1983

Appendix 3 :References

1. Prochownik, E.V., S. Panem, and W.H. Kirsten, *Type C virus induced by iododeoxyuridine in the human embryonic cell strain HEL-12.* The Journal of General Virology, 1979. **42**(2).
2. Holder, W.D., Jr., W.G. Robey, and G.F. Vande Woude, *Activation of a C-type virus from the human carcinoma cell line HBT-3 by iododeoxyuridine and testosterone.* Nature, 1974. **249**(459).
3. Stewart, S.E., et al., *Studies on the intracisternal A-type particles in mouse plasma cell tumors: induction of maturation of the particles.* Annals of the New York Academy of Sciences, 1975. **243**.
4. Bikel, I., *Properties of two distinct classes of leukaemia viruses induced from C3H mouse cells.* Virology, 1976. **69**(2).
5. Lowy, D.R., *Infectious murine leukaemia virus from DNA of virus-negative AKR mouse embryo cells.* Proceedings of the National Academy of Sciences of the United States of America, 1978. **75**(11).
6. Stephenson, J.R., J.D. Crow, and S.A. Aaronson, *Differential activation of biologically distinguishable endogenous mouse type C RNA viruses: interaction with host cell regulatory factors.* Virology, 1974. **61**(2).
7. Stephenson, J.R. and S.A. Aaronson, *Induction of an endogenous B-tropic type C RNA virus from SWR/J mouse embryo cells in tissue culture.* Virology, 1976. **70**(2).
8. Moll, B., J.W. Hartley, and W.P. Rowe, *Induction of B-tropic and N-tropic murine leukaemia virus from B10.BR/SgLi mouse embryo cell lines by 5-iodo-2'-deoxyuridine.* Journal of the National Cancer Institute, 1979. **63**(1).
9. Khan, A.S., J. Muller, and J.F. Sears, *Early detection of endogenous retroviruses in chemically induced mouse cells.* Virus Research, 2001. **79**(1-2).
10. Wu, A.M., et al., *A survey on the effect of steroid hormone on type C virus production from cultured murine cells.* Cancer Research, 1976. **36**(6).
11. Takeuchi, T., et al., *Characterization of mouse 3T3-swiss albino cells available in Japan: necessity of quality control when used as feeders.* Japanese Journal of Infectious Diseases, 2008. **61**(1).
12. Paran, M., et al., *Adrenal corticosteroids enhance production of type-C virus induced by 5-iodo-2'-deoxyuridine from cultured mouse fibroblasts.* Proceedings of the National Academy of Sciences of the United States of America, 1973. **70**(8).
13. Dunn, C.Y., *Interactions of chemical inducers and steroid enhancers of endogenous mouse type-C RNA viruses.* Virology, 1975. **66**(2).
14. Aaronson, S.A. and C.Y. Dunn, *High-frequency C-type virus induction by inhibitors of protein synthesis.* Science (New York, N.Y.), 1974. **183**(123).
15. Greenberger, J.S. and S.A. Aaronson, *Cycloheximide induction of xenotropic type C virus from synchronized mouse cells: metabolic requirements for virus activation*, 1975.
16. Yoshikura, H. and F. Zajdela, *[Induction of endogenous C-type virus by 5-iododeoxyuridine and augmentation of this induction by polycyclic hydrocarbons].* Comptes Rendus Hebdomadaires Des Séances De l'Académie Des Sciences. Série D: Sciences Naturelles, 1976. **282**(14).
17. Yoshikura, H., *Clonal Variation of the Mouse Cells in the Endogenous C-type Virus Induction by 5-iodo-2'-deoxyuridine.* Journal of General Virology, 1975. **28**(2): p. 179 -184
18. Kuff, E.L., et al., *Differential response of type C and intracisternal type A particle markers in cells treated with iododeoxyuridine and dexamethasone.* Journal of Virology, 1976. **19**(2).
19. Lowy, D.R., et al., *Murine leukaemia virus: high-frequency activation in vitro by 5-iododeoxyuridine and 5-bromodeoxyuridine.* Science (New York, N.Y.), 1971. **174**(5).
20. Teich, N., et al., *Studies of the mechanism of induction of infectious murine leukaemia virus from AKR mouse embryo cell lines by 5-iododeoxyuridine and 5-bromodeoxyuridine.* Virology, 1973. **51**(1).
21. Tennant, R.W., et al., *Induction of Retrovirus Gene Expression in Mouse Cells by Some Chemical Mutagens.* Cancer Research, 1982. **42**(8).
22. Rascati, R.J. and R.W. Tennant, *Induction of endogenous murine retroviurs by hydroxyurea and related compounds.* Virology, 1978. **87**(1).
23. Rapp, U.R. and R.C. Nowinski, *Endogenous ecotropic mouse type C viruses deficient in replication and production of XC plaques.* Journal of Virology, 1976. **18**(2): p. 411-417.
24. Stephenson, J.R. and S.A. Aaronson, *Genetic factors influencing C-type RNA virus induction.* The Journal of Experimental Medicine, 1972. **136**(1).

25. Ihle, J.N., et al., *Effect of glucocorticoids on activation of leukaemia virus in AKR mouse embryo cells.* Cancer Research, 1975. **35**(2).
26. Chattopadhyay, S.K., et al., *Correlation of the induction of transcription of the AKR mouse genome 5-lododeoxyuridine with the activation of an endogenous murine leukaemia virus.* Cancer Research, 1979. **39**(5).
27. Rapp, U.R., *Kinetics of expression of infectious ecotropic, xenotropic, and mink cell focus-forming murine leukaemia virus after 5-iododeoxyuridine induction of cells from high- and low-leukaemia mouse strains.* Journal of Virology, 1983. **45**(2).

Appendix Table 4 Review of studies utilising 5-AzaC for the induction of MLV

Table 4 Appendix

Inducer	*Cell line*	*Other chemicals*	*Released virus*	*Dose*	*Start Peak End*	*Observations*	*Detection*	*Ref*
5-AZAC	AKR clone 26A	Cytidine analogs: 5-aza-2'-deoxy cytdine (5-aza-dC), (5-aza), 5,6-dihydro-5-aza cytidine (5, 6- diH -5-aza) and (6-aza).	Endogenous murine retrovirus	Multiple	both without subculture (P0) and with subculture (P1).	dose dependent effect 5-aza-dC (0.06 ~M) > 5-aza (0.3 ~M) > 5,6-diH-5-aza (15-30 /~M) > 6-aza (/ >150~M).	n/a	[1] 1988
	K-BALB AKR-2B	Bromodeoxyuridine Cycloheximide	Ecotropic endogenous MLV and xenotropic endogenous MLV	2 µg/ml for 5-azaC, 30 µg/ml for BrUrd, or 20 µg/ml for cyclo	Peak at 10h End at 40h	Optimal : 2,µg/ml (8.1 M) 5-azaC and 20 µg/ml (65 µM). BrUrd (comparable)	Ecotropic tested on NIH and xenotropic on NRK cells	[2] 1975
	K-BALB	-	IAP	24 hr 5-aza 0.50-2µg/ml+ 20 µg/ml IdU	24-hr treatment n/a	Concentration and time dependent.	-n/a	[3] 1983
	K-BALB	alone and +IdU	C type and IAP	5-IDU 30 µg/ml for 24h 5-AzaC 2 µg/ml	start Day1 peak Day2 n/a	n/a	PERT TEM Infection of NIH3T3	[4] 2001
	AKR BALB/c NIH/ 3T3	(BP,) benzo (a)-pyrene ; (BPD), 7.8-diol-benzo (a) pyrene; (BPDE), 7.8-diol-9, 10-epoxy-benzo(a) pyrene; (5-IDU)	Endogenous RVs	5-IDU from 20 to 50 µg/ml (BP) at a 100 mM 5-AzaC 2-4ug/ml	Start :Day 2 and day 3 day 4(AZA) Declined by day6	5- azacytidine induced >90% of AKR cells at a concentration of 2 to 4 µg/ml	Mainly comparison of 5-IDU and others	[5] 1982
	C3H 10T½	5-azacytidine (5-azaC)	Endogenous MLV related to mo-MLV and IAP	*24h 5-AzaC 3 µM*	increase after passage 5	mo-MLV LTR sequences and env MLV in DNA	Northern and souther hybridization	(6) 1986

Table 4 Appendix : References

1. Rascati, R.J., *Effects of cytidine analogs on methylation of DNA and retrovirus induction.* Mutation Research, 1988. **208**(1).
2. Niwa, O., A. Declève, and H.S. Kaplan, *Potentiating effect of iododeoxyuridine on MuLV replication in mouse embryo fibroblasts.* Virology, 1975. **67**(1).
3. Lasneret, J., et al., *Activation of intracistemal a particles by 5-azacytidine in mouse Ki-BALB cell line.* Virology, 1983. **128**(2).
4. Khan, A.S., J. Muller, and J.F. Sears, *Early detection of endogenous retroviruses in chemically induced mouse cells.* Virus Research, 2001. **79**(1-2).
5. Tennant, R.W., et al., *Induction of Retrovirus Gene Expression in Mouse Cells by Some Chemical Mutagens.* Cancer Research, 1982. **42**(8).
6. Hsiao, W. L., Gattoni-Celli, S., & Weinstein, I. B. (1986). Effects of 5-azacytidine on expression of endogenous retrovirus-related sequences in C3H 10T1/2 cells. *Journal of Virology*, *57*(3), 1119–1126.

Appendix 5 Supplemented Hormonal Epithelium Medium (SHEM) :recipe

Dulbecco /F12 2:1
FBS 10%
Insulin 5µg/ml
Transferrin 5µg/ml
Adenine 0.18mM
Triodothyronine 2nmol/l
Hydrocortisone 0.4 mg/ml
Cholera toxin subunit 0.1nM
EGF 10ng/ml*
Penicillin-Streptomycin 50IU/ml

*EGF is added at 10ng/ml beginning at first feeding, 3 days after plating.

Appendix 6: Full length sequence of CF1-cell related MLV virus (FASTA format) (deposited in GenBank Acc No MG525170)

```
LOCUS       CF1-MLV                8270 bp    RNA     linear       15-NOV-2017
DEFINITION  Murine leukaemia virus isolate CF1 cell line culture.
ACCESSION   CF1-MLV
VERSION
KEYWORDS    .
SOURCE      murine leukaemia virus
  ORGANISM  murine leukaemia virus
            Unclassified.
REFERENCE   1  (bases 1 to 8270)
  AUTHORS   Kakisi,O.K.
  TITLE     RESCUE OF MURINE RETROVIRUSES FROM HUMAN LIMBAL STEM CELLS
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 8270)
  AUTHORS   Kakisi,O.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2017) Infectious Diseases, Imperial College
            London, Wright-Flemming Institute, Praed street, London W2 1PG, UK
COMMENT     Bankit Comment: ALT EMAIL:okakisi@gmail.com.
            Bankit Comment: TOTAL # OF SEQS:1.

            ##Assembly-Data-START##
            Assembly Method       :: Geneious v. R8
            Coverage              :: 94
            Sequencing Technology :: Illumina
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..8270
                     /organism="murine leukaemia virus"
                     /mol_type="other RNA"
                     /isolate="CF1 cell line culture"
                     /specimen_voucher="59325428"
                     /cell_line="MEF:CF1 ATCC SCRC-1040"
                     /lab_host="Imperial College London"
                     /country="United Kingdom"
                     /collection_date="OCT 2014"
                     /collected_by="Dr Ourania Kakisi"
     repeat_region   1..56
                     /rpt_family="5-LTR R"
                     /rpt_type=long_terminal_repeat
                     /rpt_unit_range=1..56
```

repeat_region 57..130
/rpt_family="U5 repeat"
/rpt_type=long_terminal_repeat
/rpt_unit_range=57..130
gene 363..2240
/gene="glyco-gag"
CDS 363..2240
/gene="glyco-gag"
/codon_start=1
/product="glyco-gag"
/translation="MGDVSEASGARWVAQSVSPSPDRFGLFGAPPLSEGYVVLLGDER
SKPSPPPSEFLLSVFRRNRAARLVCLSIVLSFVCSLLFWTASKNMGQTVTTPLSLTLE
HWEDVQRIASNQSVDVKKRRWVTFCSAEWPTFGVGWPQDGTFNLDIILQVKSKVFSPG
PHGHPDQVPYIVTWEAIAYEPPPWVKPFVSPKLSPSPTAPILPSGPSTQPPPRSALYP
ALTPSIKPRPSKPQVLSDNGGPLIDLLSEDPPPYGGQGLSSSDGDGDREEATSTSEIP
APSPIVSRLRGKRDPPAADSTTSRAFPLRLGGNGQLQYWPFSSSDLYNWKNNNPSFSE
DPGKLTALIESVLTTHQPTWDDCQQLLGTLLTGEEKQRVLLEARKAVRGNDGRPTQLP
NEVDAAFPLERPDWDYTTQRGRNNLVLYRQLLLAGLQNAGRSPTNLAKVKGITQGPNE
SPSAFLERLKEAYRRYTPYDPEDPGQETNVSMSFIWQSAPDIGRKLERLEDLKSKTLG
DLVREAERIFNKRETPEEREERVRRETEEKEERRRAEEEQKEKERDRRRHREMSKLLA
TVVSGQRQDRQGGERRRPQLDKDQCAYCKEKGHWAKDCPKKPRGPRGPRPQTSLLTLD
D"
gene 627..2240
/gene="gag"
CDS 627..2240
/gene="gag"
/codon_start=1
/product="polyprotein"
/translation="MGQTVTTPLSLTLEHWEDVQRIASNQSVDVKKRRWVTFCSAEWP
TFGVGWPQDGTFNLDIILQVKSKVFSPGPHGHPDQVPYIVTWEAIAYEPPPWVKPFVS
PKLSPSPTAPILPSGPSTQPPPRSALYPALTPSIKPRPSKPQVLSDNGGPLIDLLSED
PPPYGGQGLSSSDGDGDREEATSTSEIPAPSPIVSRLRGKRDPPAADSTTSRAFPLRL
GGNGQLQYWPFSSSDLYNWKNNNPSFSEDPGKLTALIESVLTTHQPTWDDCQQLLGTL
LTGEEKQRVLLEARKAVRGNDGRPTQLPNEVDAAFPLERPDWDYTTQRGRNNLVLYRQ
LLLAGLQNAGRSPTNLAKVKGITQGPNESPSAFLERLKEAYRRYTPYDPEDPGQETNV
SMSFIWQSAPDIGRKLERLEDLKSKTLGDLVREAERIFNKRETPEEREERVRRETEEK
EERRRAEEEQKEKERDRRRHREMSKLLATVVSGQRQDRQGGERRRPQLDKDQCAYCKE
KGHWAKDCPKKPRGPRGPRPQTSLLTLDD"
misc_difference 1608..1609
/gene="gag"
/note="intolerant N415H in MHR"
gene 5771..7780
/gene="envelope"
CDS 5771..7780
/gene="envelope"
/codon_start=1

/product="envelope"
/translation="MESTTLSKPFKNQVNPWGPLIVLLILGGVNPVTLENSPHQVFNL
TWEVTNGDRETVWAITGNHPLWTWWPDLTPDLCMLALHGPSYWGLEYRAPFSPPPGPP
CCSGSSDSTPGCSRDCEEPLTSYTPRCNTAWNRLKLSKVTHAHNGGFYVCPGPHRPRW
ARSCGGPESFYCASWGCETTGRASWKPSSSWDYITVSNNLTSDQATPVCKGNEWCNSL
TIRFTSFGKQATSWVTGHWWGLRLYVSGHDPGLIFGIRLKITDSGPRVPIGPNPVLSD
RRPPSRPRPTRSPPPSNSTPTETPLTLPEPPPAGVENRLLNLVKGAYQALNLTSPDKT
QECWLCLVSGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSEVTGQGLCIGAVPK
THQVLCNTTQKTSDGSYYLAAPTGTTWACSTGLTPCISTTILDLTTDYCVLVELWPRV
TYHSPSYVYHQFERRAKYKREPVSLTLALLLGGLTMGGIAAGVGTGTTALVATQQFQQ
LQAAMHDDLKEVEKSITNLEKSLTSLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFY
ADHTGLVRDSMAKLRERLSQRQKLFESQQGWFEGLFNKSPWFTTLISTIMGPLIILLL
ILLFGPCILNRLVQFIKDRISVVQALVLTQQYHQLKTIEDCKSRE"
misc_feature 5771..5787
/gene="envelope"
/note="splice donor"
repeat_region 8196..8263
/rpt_family="R element"
/rpt_type=long_terminal_repeat
/rpt_unit_range=8196..8263
BASE COUNT 2111 a 2386 c 2017 g 1756 t
ORIGIN
1 cgatagactg agtcgcccgg gtacccgtgt atccaataaa gccttttgct gttgcatccg
61 aatcgtggtc tcgctgatcc ttgggagggt ctcctcagag tgattgactg cccagcctgg
121 gggtctttca tttgggggct cgtccgggat ttggagaccc ccgcccaggg accaccgacc
181 caccgtcggg aggtaagctg gccagcgatc gttttgtctc cgtctctgtc tttgtgcgtg
241 tgtgtgtgtg tgccggcatc tacttttttgc gcctgcgtct gattctgtac tagttagcta
301 actagatctg tatctggcgg ctccgtggaa gaactgacga gttcgtattc ccgaccgcag
361 ccctgggaga cgtctcagag gcatcggggg cccgctgggt ggcccaatca gtaagtccga
421 gtcctgaccg attcggacta tttggggccc ctcctttgtc ggaggggtac gtggttcttt
481 taggagacga gaggtccaag ccctcgccgc ctccatctga atttttgctt tcggtttttc
541 gccgaaaccg cgccgcgcgt cttgtctgtc tcagtattgt tttgtcattt gtctgttcgt
601 tattgttttg gaccgcttct aaaaacatgg gacagaccgt aaccacccct ctgagtctga
661 ccctagaaca ctgggaagat gtccagcgca tcgcgtccaa tcagtccgta gatgtcaaga
721 agagacgctg ggtcaccttc tgctctgccg agtggccaac tttcggtgta gggtggccac
781 aagatggtac ttttaatttg gacattattc tacaggttaa atctaaggtg ttctctcctg
841 gtccccacgg acacccggat caggtcccat atattgtcac ctgggaggct attgcctatg
901 aacccccctcc gtgggtcaaa ccttttgtct ctcccaaact ctccccctct ccaaccgctc
961 ccatcctccc atccggtcct tcgacccaac ctccgccccg atctgccctt taccctgctc
1021 ttacccccctc tataaaaccc agaccttcta aacctcaggt tctctccgat aatggcggac
1081 ctctcattga ccttctctca gaagaccctc cgccgtacgg aggacaggga ctgtcctcct
1141 ctgacggaga tggcgacaga gaagaggcca cctccacttc tgagattcct gccccctctc
1201 ccatagtgtc tcgcctgcgg ggcaaaagag acccccccgc ggcagattcc accacctctc
1261 gggctttccc actccgtttg gggggtaatg gtcagttgca gtactggccg ttttcctcct
1321 ctgatctata taactggaaa aataataatc cttccttctc tgaggatcca ggtaaactga
1381 ctgcattgat tgaatccgtc ctcaccaccc accagcccac ctgggatgat tgccagcaat

1441 tattagggac tctgcttacc ggggaggaga agcagcgggt gctcctggaa gcccgaaagg
1501 ctgtccgggg caacgatggg cgccccaccc aactgcccaa cgaggttgac gctgcttttc
1561 cccttgaacg tcccgattgg gattacacca cccaaagagg taggaacaac ctagttctct
1621 atcgccagtt gcttttagca ggtctccaaa acgcgggccg aagccccacc aatttggcca
1681 aggtaaaagg aataacccag ggacctaatg agtccccctc agccttccta gagagactca
1741 aggaagccta tcgcagatac actccttatg atcctgagga ccctgggcaa gaaacgaatg
1801 tatctatgtc attcatctgg cagtccgctc cagacattgg tcgaaagtta gagcggttag
1861 aagacttaaa aagtaaaact ttaggagatt tagtgagaga agccgaaagg atctttaata
1921 aaagagaaac tccagaagaa agagaagagc gcgttagaag ggagacagag gaaaaggaag
1981 agcggcgtag ggcagaagag gagcagaaag agaaagagag ggaccgtagg agacatagag
2041 aaatgagcaa acttttggcc accgtagtta gtggacagag acaggataga caggggggag
2101 agcgaaggag gccccaactt gacaaggatc aatgtgccta ctgtaaagaa aaaggacact
2161 gggctaaaga ctgcccaaag aagccacggg gtccccgagg accacgaccc cagacctccc
2221 tcctgacttt agacgactag gggggtcagg gtcaggagcc ccccccgaa cccaggataa
2281 ccctcactgt cgggggggcaa cccgtcacct tcctggtgga tactggggcc caacactccg
2341 tgctgactca aaatcctggg cccctaagtg acaggtcggc ttgggtccaa ggggctactg
2401 gaggaaagcg gtatcgctgg accacagatc gcaaggtaca cctagctacc ggtaaggtca
2461 ctcactcttt cctccatgtg ccagactgcc cctacccctt gctaggaaga gacttgttaa
2521 ctaaactaaa agcccagatc cactttgagg gatcaggagc ccaggttgtg ggaccaaaag
2581 gacagcctct acaggtgttg accctaaacc tagaagatga gtatcggctt tatgagacct
2641 cagcagagcc ggaagcttct cctgggtcca cctggctctc cgactttccc caggcctggg
2701 cagaaaccgg gggcatggga ctggcagttc gccaggcgcc tctaatcata cctctgaagg
2761 caacctccac ccctgtgtcc ataaaacaat accccatgtc acaggaagcc agactgggga
2821 tcaagcccca catacagagg ctgttggacc agggaatact ggtaccctgc cagtcccct
2881 ggaatacacc cctgctaccc gttaagaaac caggaactaa cgattatagg cctgtccagg
2941 atctgagaga agtcaacaag cgggtggaag acatccaccc caccgtgccc aatccttata
3001 acctcctaag tgggctccca ccgtcccacc ggtggtatac tgtgcttgat ttaaaggatg
3061 ccttttctg cctgagactc caccccacca gtcagcctct cttcgccttc gagtggagag
3121 atccagggat gggaatctca ggacaattaa cctggaccag actcccacag ggtttcaaaa
3181 acagtcccac cctgtttgat gaggcactac acagagacct agcagacttc cggatccagc
3241 acccagactt gatcctgcta cagtacgtgg atgacatact actggccgcc acttctgagc
3301 tcgactgcca acaaggtact cgggccctgt tactaaccct aggaaacctc gggtatcggg
3361 cctcggccaa gaaagcccaa ctttgccaga aacaggtcaa gtatctgggg tatctcctaa
3421 aagagggtca gagatggctg actgaggcca gaaaagagac tgtgatgggg cagcctactc
3481 cgaagacccc tcgacaacta agggagttcc tagggacggc aggcttctgt cgcctctgga
3541 tccctgggtt tgcggaaatg gcagcccccct tgtatcctct taccaaaacg gggactctgt
3601 tcaattgggg cccagaccag caaaaggcct atcaagaaat caaacaggcc ctcctaactg
3661 cccccgccct gggattgcca gatttgacta agccctttga actctttgtc gacgagaagc
3721 agggctacgc caaaggcgtc ctaacgcaaa aactgggacc ttggcgtcgg cccgtggcct
3781 acctgtccaa aaagctagat ccagtggcag ctgggtggcc cccttgtcta cggatggtag
3841 cagccattgc cgttctgaca aaagatgcag gcaagctaac tatgggacag ccgctagtca
3901 tcctggcccc ccatgcagta gaggcactgg tcaagcaacc ccctgaccgc tggctatcca
3961 acgcccgcat gacccactac caggcaatgc tcctagacac tgaccgagtt cagttcggac
4021 cagtggtggc cctcaatcct gccaccttac tccctctccc ggaagaagga gccccccatg
4081 attgcctcga gatcttggct gaaacgcatg gaaccagacc ggatctcacc gaccagccca
4141 tcccagacgc cgaccacacc tggtataccg atgggagcag ctttttgcaa gaaggacagc

4201 gaaaggctgg ggcagcggtg accactgaga ccgaggtaat ctgggcaagg gcactgccgg
4261 ctggaacatc tgcccaacgg gccgaactga tagcactcac tcaagccttg aagatggcag
4321 aaggtaagag gctaaacgtt tacactgaca gccgatatgc tttcgccacg gcccatatcc
4381 atggagagat ctataggagg cgagggttgc taacctcaga gggtagagaa atcaaaaaca
4441 agagcgagat cctggcttta ctgaaagctc tttttctgcc aaagagactc agtataatcc
4501 actgcctggg gcatcaaaaa ggagatagtg ccgaggctag gggcaaccgc ctagcagacc
4561 aagcggcccg ggaggcagcc ataaagacgc ctccagatac atccactctc cttatagagg
4621 attcaacccc atatacgcct gcctatttcc attatactga aacagatcta aagaaactca
4681 gagagcttgg ggccacctat aatcagagca aaggatactg ggtcttccaa ggcaagccgg
4741 tgatgcccga ccaatttgta tttgaactgt tagactcact ccaccggctc acccacctcg
4801 gctaccagaa aatgaaagca ctccttgaca gaggagaaag cccctactac atgctaaacc
4861 gggacaaaac cctccaatat gtggcagatt cctgcacggt ctgtgcccaa gtaaatgcca
4921 gcaaagctaa aatcggggca ggagtgagag tacgaggaca tcggccaggc tcccattggg
4981 agatcgattt tacagaagtc aagccagggc tgtatgggta caagtacctc ctggtattcg
5041 tggacacctt ctctggttgg gtggaagcct ttccaaccaa gagagaaaca gcaagagtcg
5101 tgtccaagaa attgctggaa gaaatattcc cgagattcgg aatgccacag gtattgggat
5161 ctgataacgg gcctgccttc acctcccagg taagtcagtc ggtggccgat ttactgggga
5221 tcgattggaa attacattgt gcttatagac cccagagttc aggtcaggta gaaagaatga
5281 atagaaccat caaggagact ctaactaaat taacgcttgc agctggcact agagactggg
5341 tactcctact ccccttagct ctttaccgag cccggaacac tccgggcccc catggactga
5401 ctccgtatga aatcttgtac ggggcgcccc cgccccttgt taacttccat gaccccgaca
5461 tgtcagaatt aactaatagc ccatctctcc aagctcactt acaggccctc caaacggtgc
5521 agcgagaaat ttggaaacca ctggccgagg cctaccggga ccaactagac caaccagtga
5581 taccacaccc cttccggatt ggagactccg tgtgggtgcg ccggcaccag accaaaaact
5641 tagaacctcg ctggaaggga ccctacaccg tcctactgac cacccccacc gctctcaagg
5701 tagacggcat ctctgcatgg atacacgccg cccacgtcaa ggcagcgacc acgccccga
5761 taaaaccatc atggagagta caacgctctc aaaaacccttt aaaaatcagg ttaacccgtg
5821 gggcccccta attgtccttc tgattctcgg aggagtcaac cccgttacgt tggaaaacag
5881 cccccaccag gtttttaacc tcacctggga agtgactaat ggagaccgag aaacggtgtg
5941 ggcaataacc ggcaatcacc ctctgtggac ttggtggcct gacctcacac cagatctctg
6001 tatgttggcc ctccacgggc cgtcctattg gggcctagaa tatcgggctc ctttttctcc
6061 tcccccgggg ccccccctgct gttcaggaag cagcgactcc acgccaggct gttccagaga
6121 ttgtgaggag cccctgactt catatactcc ccggtgcaat acggcctgga acagacttaa
6181 gttatctaaa gtgacacatg cacacaatgg aggattctat gtctgccccg ggccacatcg
6241 cccccggtgg gcccggtcat gtggtggtcc agaatccttc tattgtgcct cttggggctg
6301 cgaaaccaca ggccgagcat cctggaaacc atcctcgtcc tgggactaca tcacagtaag
6361 caacaatcta acctcagacc aggcaacccc agtatgcaaa ggtaatgagt ggtgcaactc
6421 cttaactatc cggttcacga gctttggaaa acaggccacc tcctgggtca caggccattg
6481 gtggggattg cgcctatacg tctctggaca tgacccaggg ctcatctttg ggatccgact
6541 taaaattaca gactcggggc cccgggtccc aatagggcca aaccccgtct tgtcagaccg
6601 acgaccacct tcccggccta gacccaccag atctcccccg ccttcaaact ccaccccaac
6661 cgagacaccc ctcaccctcc ccgaaccccc gccagcggga gtcgaaaacc gattgttaaa
6721 tctagtaaaa ggagcctacc aagccctcaa cctcaccagt cctgataaaa cccaagagtg
6781 ctggttatgc ctagtatcgg gacccccata ctacgagggg gttgccgtcc taggtaccta
6841 ctccaaccat acttctgccc cagctaactg ctctgtggcc tctcaacaca aattgacctt
6901 gtccgaagta accggacagg gactctgcat aggagcggtc cctaaaaccc atcaagtctt

6961 gtgtaatacc acccaaaaga caagcgatgg gtcctactat ttggccgctc ccacaggaac
7021 tacctgggct tgtagtactg gacttactcc ctgtatctca accaccatac ttgacctcac
7081 caccgattac tgtgttctgg tcgagctttg gccaagggtg acctaccatt cccctagtta
7141 tgtttaccac caatttgaaa gacgagccaa atataaaaga gaacccgtct cactaactct
7201 ggccctacta ttaggaggac tcactatggg cggaattgcc gctggagtgg gaacagggac
7261 taccgcccta gtggccactc agcagttcca acaactccag gctgccatgc acgatgacct
7321 taaagaagtt gaaaagtcca tcactaatct agaaaaatct ttgacctcct tgtccgaagt
7381 agtgttacag aatcgtagag gcctagatct actattccta aaagagggag gtttgtgtgc
7441 tgccttaaaa gaagaatgct gtttctatgc cgaccacaca ggattggtac gggatagcat
7501 ggccaaactt agagaaagat tgagtcagag acaaaagctc tttgaatccc aacaagggtg
7561 gtttgaaggg ctgtttaata agtccccttg gttcaccacc ctgatatcca ccatcatggg
7621 tcccctgata atcctcttgt taattttact ctttgggcct tgtattctca atcgcctggt
7681 ccagtttatc aaagacagga tttcggtagt gcaggccctg gttctgactc aacaatatca
7741 tcaacttaag acaatagaag attgtaaatc acgtgaataa aagattttat tcagtttaca
7801 gaaagagggg ggaatgaaag acccccttcat aaggcttagc cagctaactg cagtaacgcc
7861 attttgcaag gcatgggaaa ataccagagc tgatgttctc agaaaaacaa gaacaaggaa
7921 gtacagagag gctggaaagt accgggacta gggccaaaca ggatatctgt ggtcaagcac
7981 tagggccccg gcccagggcc aagaacagat ggtccccaga aatagctaaa acaacaacag
8041 tttcaagaga cccagaaact gtctcaaggt tccccagatg accggggatc aaccccaagc
8101 ctcatttaaa ctaaccaatc agctcgcttc tcgcttctgt acccgcgctt attgctgccc
8161 agctctataa aaagggtaag aaccccacac tcggcgcgcc agtcctccga tagactgagt
8221 cgcccgggta cccgtgtatc caataaagcc ttttgctgtt gcatccgaaa

Appendix 7 The chromosomal locations of 237 XMRV integrations in 22rv1 cells (32 sequences deposited in GenBank Acc Not MG525138-69*)

XMRV integrations in 22rv1 cells

chr1	21,288,500
chr1	23,880,701
chr1	25,566,829
chr1	26,250,636
chr1	32,574,640
chr1	41,181,766
chr1	41,941,914
chr1	43,486,675
chr1	52,499,446
chr1	53,612,536
chr1	65,978,784
chr1	68,111,114
chr1	82,445,016
chr1	118,661,589
chr1	145,384,664
chr1	145,489,474
chr1	148,889,720
chr1	149,992,944
chr1	150,208,006
chr1	150,252,463
chr1	150,882,515
chr1	151,024,438
chr1	154,307,162
chr1	156,736,580
chr1	167,896,912
chr1	181,322,516
chr1	183,439,901
chr1	192,469,579
chr1	199,996,944
chr1	217,180,405

chr1	222,409,262
chr1	226,891,929
chr2	16,508,419
chr2	70,352,048
chr2	75,873,614
chr2	85,186,626
chr2	86,426,861
chr2	155,341,568
chr2	180,138,303
chr2	191,354,401
chr2	193,093,310
chr2	198,478,168
chr2	212,426,165
chr2	219,262,799
chr2	234,180,208
chr2	238,391,587
chr3	10,234,408
chr3	47,553,621
chr3	48,515,105
chr3	50,640,627
chr3	101,436,298
chr3	120,949,856
chr3	122,347,763
chr3	127,318,014
chr3	128,067,572
chr3	130,698,158
chr3	141,659,537
chr3	148,323,174
chr3	169,490,500
chr3	186,501,628
chr4	986,311
chr4	8,652,728
chr4	10,352,665
chr4	15,771,730
chr4	40,194,130

chr4	100,878,617
chr4	107,958,654
chr4	130,770,633
chr4	130,770,609
chr4	185,748,458
chr5	101,553,123
chr5	114,511,117
chr5	144,336,585
chr5	147,217,064
chr5	147,217,096
chr5	153,467,778
chr5	180,287,420
chr6	10,100,507
chr6	12,344,101
chr6	30,767,245
chr6	30,851,809
chr6	31,154,917
chr6	33,699,876
chr6	34,393,131
chr6	34,522,368
chr6	35,714,202
chr6	36,409,914
chr6	46,140,403
chr6	47,430,123
chr6	64,271,552
chr6	109,533,250
chr6	131,013,720
chr6	142,442,892
chr6	157,515,586
chr7	2,233,257
chr7	18,102,048
chr7	24,182,255
chr7	25,472,658
chr7	73,795,440
chr7	90,031,995

chr7	99,031,909
chr7	99,068,775
chr7	100,875,048
chr7	150,065,149
chr8	9,007,286
chr8	21,998,992
chr8	23,482,152
chr8	38,182,999
chr8	48,439,017
chr8	50,823,071
chr8	67,974,869
chr8	81,468,370
chr8	81,489,606
chr8	82,894,873
chr8	88,265,957
chr8	100,360,063
chr8	128,028,685
chr8	128,110,114
chr8	128,076,465
chr9	13,695,895
chr9	71,735,233
chr9	76,282,731
chr9	94,921,662
chr9	111,572,465
chr9	131,902,607
chr9	132,457,818
chr9	135,231,218
chr10	5,590,455
chr10	5,591,442
chr10	5,710,909
chr10	29,399,632
chr10	63,798,122
chr10	79,259,526
chr10	98,459,292
chr10	100,757,111
chr10	105,111,493

chr11	12,239,682
chr11	15,166,029
chr11	39,753,110
chr11	44,161,800
chr11	58,791,260
chr11	64,887,882
chr11	67,034,634
chr11	67,273,466
chr11	69,452,184
chr11	69,454,505
chr11	111,232,519
chr11	111,232,483
chr11	114,046,535
chr11	126,152,492
chr12	6,921,387
chr12	8,849,751
chr12	12,536,402
chr12	16,701,042
chr12	46,844,529
chr12	51,626,427
chr12	52,418,523
chr12	71,002,307
chr12	85,767,110
chr12	91,381,876
chr12	120,836,275
chr12	110,115,203
chr12	115,108,616
chr12	119,005,445
chr12	124,672,769
chr13	30,997,438
chr13	63,584,841
chr13	99,850,190
chr13	30,055,448
chr13	38,036,170
chr13	38,037,992
chr14	38,050,816

chr14	38,064,913
chr14	41,489,239
chr14	55,002,711
chr14	57,578,329
chr14	68,185,910
chr14	69,865,715
chr14	74,254,818
chr14	88,636,452
chr14	102,605,023
chr15	40,360,034
chr15	43,398,414
chr15	66,732,886
chr15	85,290,889
chr15	93,353,208
chr16	30,546,614
chr16	30,753,157
chr16	46,533,322
chr16	50,070,588
chr16	53,798,971
chr16	80,868,995
chr16	89,741,833
chr17	7,960,812
chr17	8,041,565
chr17	8,056,384
chr17	16,200,864
chr17	18,162,490
chr17	19,533,586
chr17	27,170,167
chr17	73,782,198
chr17	73,873,834
chr17	80,290,396
chr18	658,996
chr18	61,350,693
chr18	76,971,862
chr19	4,993,319
chr19	13,273,274

chr19	36,628,372
chr19	39,328,603
chr19	44,174,517
chr19	48,891,273
chr19	49,108,338
chr19	658,968
chr19	796,496
chr19	10,528,824
chr19	49,108,336
chr19	49,930,421
chr20	6,103,399
chr19	26,213,080
chr19	30,160,203
chr19	56,339,696
chr19	56,516,671
chr19	31,781,780
chr19	61,464,478
chr19	61,467,089
chr21	38,444,597
chr22	23,879,956
chr22	38,413,827
chr22	45,798,782
chr22	47,573,582
chrX	53,341,017
chrX	128,811,224
chrX	66,745,678
chrX	128,796,177
chrX	135,060,559
chrX	149,501,224

The sequences filled in GRAY are deposited in GenBank (See Appendix 8). These were considered true as having >50 reads in the sample pool of circa 850,000 for XMRV integrations in 22rv1 cells.

Appendix 8 Example GenBank submission for integration site sequences for XMRV in 22rv1 cells (Acc Not MG525138-69)

```
LOCUS       Seq1                149 bp    DNA     linear   PRI 14-NOV-2017
DEFINITION  human 22rv1 cell line genomic integration site of 5'LTR XMRV :
            chromosome 1;AL021154.1; p36.12.
ACCESSION   Seq1
VERSION
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149)
  AUTHORS   Kakisi,O.K.
  TITLE     RESCUE OF MURINE RETROVIRUSES FROM HUMAN LIMBAL STEM CELLS
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 149)
  AUTHORS   Kakisi,O.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2017) Infectious Diseases, Imperial College
            London, Wright-Flemming Institute, Praed street, London W2 1PG, UK
COMMENT     Bankit Comment: ALT EMAIL:okakisi@gmail.com.
            Bankit Comment: TOTAL # OF SEQS:32.
                ##Assembly-Data-START##
            Assembly Method       :: Geneious v. R8
            Sequencing Technology :: Illumina
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..149
                     /organism="Homo sapiens"
                     /proviral
                     /mol_type="genomic DNA"
                     /isolation_source="cell culture"
                     /db_xref="taxon:9606"
                     /cell_line="22rv1 (ATCC CRL-2505)"
                     /lab_host="Wright Fleming Institute- Imperial College London"
                     /country="United Kingdom"
                     /collection_date="Nov 2014
     misc_recomb     119
                     /note="integration site"
ORIGIN
        1 gaccgtggct gaatgagact ggtgtcgaca ctagtgggat cggggtagtg agacaccttg
       61 agaccgggac tttgggggtt gcctttgaca aaccagatta aggagggagg agttctctgt
      121 gaaagacccc accataaggc ttagcacgc //
```

Appendix Table 9
Viral LTR sequences used in further mapping of integration sites of MLVs via NGS

Table Appendix 9

Viral LTRs	name	Acc No	interest	5' LTR sequence	direction
MLV	RFB-MLV	MLU43202	primer specific U3	AATGAAAGACCCCTTCATAAGGCTTAGC CAGCTAACTGCAGTAACGCCATTTTGCA AGGCATGGA	rev
	FBR-MLV	K02712 (MSVMUSV)	one mismatch with primer U3	AATGAAAGACCCCTTCATAAGGCTTAGC CAGCTAACTGCAGTAACGCCATTTTGCA AGGCATGG*G*	rev
	emv-1	DQ366147	one mismatch with primer U3	AATGAAAGACCCCTTCATAAGGCTTAGC CAGCTAACTGCAGTAACGCCATCTTGCA AGGCATGG*G*	rev
	emv-30	KJ668269	one mismatch with primer U3	AATGAAAGACCCCTTCATAAGGCTTAGC CAGCTAACTGCAGTAACGCCATTTTGCA AGGCATGG*G*	rev
	osteosarcoma	AF033814	one mismatch with primer U3	AATGAAAGACCCCTTCATAAGGCTTAGC CAGCTAACTGCAGTAACGCCATTTTGCA AGGCATGG*G*	rev
	MCF recombinant	AF136154	no mismatches with primer U3	AATGAAAGACCCC*ACCATC*AGGCTTAGC *A*AGCTA*G*CTGCAGTAACGCCATTTTGCA AGGCATGGA	rev
	xMLV recombinant	Jern et al, 2007	one mismatch with primer U3	AATGAAAGACCCCACCATAAGGCTTAG CAAGCTAGCTGCAGTAACGCCATTTTGC AAGGCATG*C*A	rev
	pMLVs	Jern et al, 2007	two mismatches with primer U3	AATGAAAGACCCCACCATCAGGCTTAG CAAGCTAGCTGCAGTAACGCCATTTTGC AAGGCAT*AA*A	rev
XMRV	22rv1 derived XMRV	FN692043	primer specific U3	AATGAAAGACCCCACCATAAGGCTTAG CACGCTAGCTACAGTAACGCCATTTTGC AAGGCATGGAAAAGTACCAGAGCTGAG TTCTCAAAAGTTACA	rev

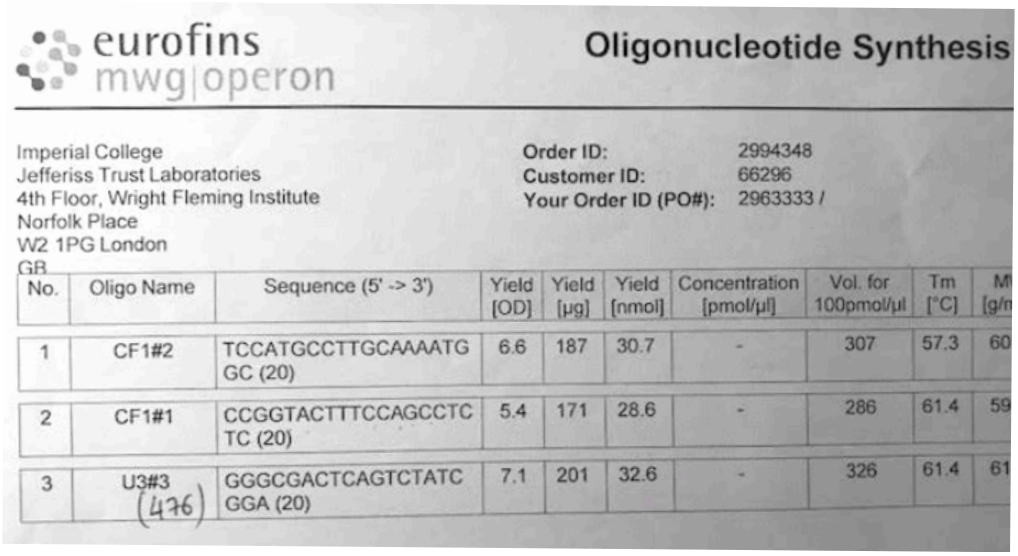
Key (Table Appendix 9)

Include first 2bp that are lost during integration (start TG)
Underlined are primer positions
In red are putative mismatches with reverse primers (located in 5' end)

Appendix 10: Lot numbers of CF1 cells used in this thesis

source	Catalog No	Lot No	where used
ATCC	SCRC-1040	59325428	Ch 3-4-6-7
	SCRC-1040.1 (IRR)	60483047	RT-PCR Ch 3-4
AMS-Bio	GSC-6001	n/a	RT-PCR Ch 3-4

Appendix 11 : Direction of U3-CF1 primers used in splinkerette PCR



eurofins mwg|operon

Oligonucleotide Synthesis

Imperial College
Jefferiss Trust Laboratories
4th Floor, Wright Fleming Institute
Norfolk Place
W2 1PG London
GB

Order ID: 2994348
Customer ID: 66296
Your Order ID (PO#): 2963333 /

No.	Oligo Name	Sequence (5' -> 3')	Yield [OD]	Yield [µg]	Yield [nmol]	Concentration [pmol/µl]	Vol. for 100pmol/µl	Tm [°C]	M [g/m
1	CF1#2	TCCATGCCTTGCAAAATG GC (20)	6.6	187	30.7	-	307	57.3	60
2	CF1#1	CCGGTACTTTCCAGCCTC TC (20)	5.4	171	28.6	-	286	61.4	59
3	U3#3 (476)	GGGCGACTCAGTCTATC GGA (20)	7.1	201	32.6	-	326	61.4	61

Appendix 12 " Copyright permission "

RE: permission

Wiley Global Permissions <permissions@wiley.com>

Fri 18/07/2014 20:05

To: Kakisi, Ourania <o.kakisi@imperial.ac.uk>;

Dear Ourania Kakisi:

Permission is hereby granted for the use requested subject to the usual acknowledgements (author, title of material, title of book/journal, ourselves as publisher). You should also duplicate the copyright notice that appears in the Wiley publication in your use of the Material.

Any third party material is expressly excluded from this permission. If any of the material you wish to use appears within our work with credit to another source, authorization from that source must be obtained.

This permission does not include the right to grant others permission to photocopy or otherwise reproduce this material except for accessible versions made by non-profit organizations serving the blind, visually impaired and other persons with print disabilities (VIPs).

Sincerely,

WILEY

Sheik Safdar | Permissions Coordinator | P: 201-748-6512 | F: 201-748-6008
John Wiley & Sons, Inc. | 111 River Street | Hoboken, NJ | 07030 | Mailstop: 4-02

From: Kakisi, Ourania [mailto:o.kakisi@imperial.ac.uk]
Sent: Friday, July 11, 2014 3:06 PM
To: Wiley Global Permissions
Subject: permission

E-theses letter template: request to reproduce my own published paper

[11 July 2014]

Dear Blackwell Publishing Ltd,

I am completing my PhD thesis at Imperial College London entitled 'Rescue of murine retroviruses in human cells'.

I seek your permission to reprint, in my thesis, the published version of a paper I published in: Transfusion Medicine, The rise and fall of XMRV Transfus Med. 2013 Jun;23(3):142-51. doi: 10.1111/tme.12049. Epub 2013 May 21. .

I would like to include the paper in the printed examination copy of my thesis and also the electronic version which will be added to Spiral, Imperial's online repository http://spiral.imperial.ac.uk/ and made available to the public under a Creative Commons Attribution-NonCommercial-NoDerivs licence.

If you are happy to grant me all the permissions requested, please return a signed copy of this letter. If you wish to grant only some of the permissions requested, please list these and then sign.

Yours sincerely,

Ourania Kakisi, MD

Permission granted for the use requested above:

I confirm that I am the copyright holder of the paper above and hereby give permission to include it in the print and electronic version of your thesis. I understand that the electronic version of the thesis will be made available, via the internet, for non-commercial purposes under the terms of the user licence.

[please edit the text above if you wish to grant more specific permission]

Signed:

Name:

Organisation:

Job title: