UNIVERSITY OF
PLYMOUTH

THE POSSIBLE ROLE OF ENDOGENOUS RETROVIRUSES IN TUMOUR DEVELOPMENT & INNATE SIGNALLING

by

EMMANUEL ATANGANA MAZE

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Biomedical Sciences

2018

COPYRIGHT STATEMENT

UNIVERSITY OF
PLYMOUTH

THE POSSIBLE ROLE OF ENDOGENOUS RETROVIRUSES IN TUMOUR DEVELOPMENT & INNATE SIGNALLING

by

EMMANUEL ATANGANA MAZE

UNIVERSITY OF PLYMOUTH

School of Biomedical Sciences

2018

"You are the light of the world. A town built on a hill cannot be hidden. Neither do people light a lamp and put it under a bowl. Instead they put it on its stand, and it gives light to everyone in the house. In the same way, let your light shine before others, that they may see your good deeds and glorify your Father in heaven."
Matthew 5:14-16

Dedicated to Mrs & Mr Carlyne &Jean-Pierre Maze Maze

AUTHOR'S DECLARATION

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

Data gathering:

The data presented in this thesis were generated by the author except as listed below:

Study on HERV-K (HML2):

Bora Agit, PhD student at Plymouth University, generated the data on freely released HERV-K (HML2) proteins (Figure 19A and B).

Shona Reeves, Undergraduate student at Bath University in 2016, generated the data for exosome-release of HERV-K (HML2) proteins (Figure 19C, D and E), supervised by the author.

Dr Sylwia Ammoun did the immunofluorescent analyses found in section 8.1 and 8.2.

Dr Emanuela Ercolano was in charge of isolation and culture of schwannomas and Schwann cells from patients at early passages.

Dr Robert Belshaw provided raw alignments of HERV-K (HML2) loci, used for the *in silico* analysis in section 8.3.

Study on PcEV:

Dr Robert Belshaw and Jack Kelly (PhD student at Plymouth University) did the bioinformatics analysis presented in section 8.4.

Claire Ham did RNA extractions from SIV-infected animals as it involved handling samples in a Biosafety Level 3 Laboratory.

Dr Neil Berry and Dr Robert Belshaw helped writing section 8.4, which has been submitted to *Journal of Virology*.

Funding:

This research degree was supported by a studentship from Plymouth University with additional support from the National Institute of Biological Standards and Control (NIBSC), Action on Hearing Loss and Action Medical Research for children.

Publications in the course of PhD degree

Published and not-linked to endogenous retroviruses:

Cellular prion protein (PrPC) in the development of Merlin-deficient tumours
Provenzano L, Ryan Y, Hilton DA, Lyons-Rimmer J, Dave F, **Maze EA**, Adams CL, Rigby-Jones R, Ammoun S, Hanemann CO.
Jul. 2017 – Oncogene

Submitted or in preparation, linked to endogenous retroviruses:

Enriched RNA-Seq analysis of the endogenous retrovirus HERVK in Mantle Cell Lymphoma cell lines: characterising a tumour associated antigen
Atkin M, Bedford F, Dickie J, Jones AP, Kiernan M, Tatkiewicz W, **Maze EA**, Agit B, Kanapin A, Belshaw R.
Submitted in 2017 – BMC Cancer

The role of human endogenous retroviral proteins in the development of Merlin-deficient tumors and as potential drug targets
Maze EA, Agit B, Reeves S, Hilton D, Lyons Rimmer J, Ercolano E, Hanemann CO, Belshaw R and Ammoun S
Manuscript in preparation

Papio Cynocephalus Endogenous Retrovirus (PcEV) expression is strongly correlated with the expression of an interferon-stimulated gene STAT1 in tissues from SIV-infected macaques
Maze EA, Ham C, Kelly J, Ussher L, Almond N, Towers G, Berry N, Belshaw R.
Submitted in 2018 – Journal of Virology

Word count of the main body: 36,878 words

Signed: 

Date: 15/11/2018

ACKNOWLEDGEMENT

This thesis would never be possible without all the people who provided me support. Those lines are to thank you all. There were a lot of you, so I apologise for the ones I may have missed.

I would like to thank Robert Belshaw, my director of study, for leading me in the field of retrovirology, for all the debates we had and for all the wisdom he shared. You have taught me a lot, I hope I did well as you first PhD student.

I would like to thank Sylwia Ammoun, my fourth supervisor, for all the trust, for always being honest and frank, and all the challenges. You contributed a lot on making me a better scientist.

I would like to thank Neil Berry, my third supervisor, for the project in NIBSC. Thanks to you I learned a lot about the SIVs and HIVs origins. Thank you for sharing the knowledge. Also, you helped me settle in South Mimms and always provide support.

I would like to thank Lynn McCallum, my second supervisor, for the help she provided. Even though you are very busy as you supervised other PhD students, you found the time to be there and provide support at key step in the PhD course.

I would like to thank Claire Adams and Paul Waines (Plymouth University lab managers) for setting up the lab and helping me for various issues I encountered; Emanuela Ercolano (Plymouth University lab technician) for the amazing in cell culture; and Claire Ham (NIBSC scientist) for setting up the experimental part of the PcEV study at NIBSC. You all made my stay and work easier. Thank you very much.

I would like to give a warm thank you to all members of the Oliver Hanemann's team. You provide a lot of support both experimentally and theoretically. More specifically, thank you Oliver (thanks for the interest you gave in our less common project, and for putting together the Derriford Research Facility. Good luck with this new cutting-edge facility); thank you Nicola and Julie (for handling all the orders and the arduous paperwork); thank you Sara and Daniele (I have learned watching you, especially the research on exosomes); thank you Jemma and Jade (for coping with the lab orders, for

all the parties and for the shDCAF1-shRNA lentiviruses); thank you Emanuela (for the lab orders and the culture of primary cells from patients).

I would like to thank all other students and research fellows I interacted with and received help from, especially Hannah Thompson, Ibrahim Halawani, Silvia Pregnolato, Connor Wood, Michele Kiernan, Garry Farnham, Liyam Labara, Robert Button, Hujaz Alqirbi, Katrina Cowan, Randa Ali, Roxane Dunbar, Victory Poloamina, Christian Good, Jon Gil, Wai Ling Kok.

A massive thank you to members of the Sylwia's crew in Oliver Hanneman's group. You are the best mates I have met through the lab, you helped me a lot to overcome experimental, theoretical and psychological issues. Thank you very much Foram Dave, Bora Agit and Lucy Provenzano.

I would like to give a big thank you to Street Factory Dance Company. You were my family in the UK, we went through and did a lot together (competitions, dancing for charities, TEDx, etc.). You empowered me for sure. Thank you lads of the seas: Toby and Jo Gorniak. Max "Silk Boogie" Revell, Sam "Sam.I.am" Holland, Ben "Benny P" Peters, B-Boy Perry Johnson, Eli Harvey, Natasha "Tasha" Jones aka B-Girl Precious, James and Sarah Matthews.

I would like to thank you Vesela, Bernard and Stephane. You are my best friends in life. You always kept on supporting me since we met back in undergraduate. I am looking forward to seeing you again.

I would like to thank my beloved Tiphany, you have always been with me since the beginning of my British adventure. You make me, not only a better scientist, but also a better man.

I would like to finally thank the pillars in my life, the Maze: Davis, Pierre, Janella, Guy Josué, Carlyne and Jean-Pierre.

Merci à tous !

1 ABSTRACT

Endogenous retroviruses (ERVs) are fossils of ancient retroviral infection in the germline. In primates they represent around 5% of the genome sequence. During time spent in the genome, being transmitted in a Mendelian fashion, copies of ERVs have accumulated mutations, which rendered them inactive. However, some of them (the most recently integrated ones) are still able to transcribe and produce viral proteins, although few are capable of re-infection. In the past often considered as unharmful 'junk DNA', recent evidence link ERVs with cancer and several inflammatory diseases. For example, a few reports demonstrate that ERVs are involved in tumour development using shRNA knock-down and over-expression systems, and their overexpression tends to correlate with inflammation status, generating the hypothesis that they can act as pathogen-associated molecular patterns (PAMPs) and bind to innate sensors.

Focusing on the Human (*Homo sapiens*) and the rhesus macaque (*Macaca mulatta*), the main aims of this thesis are to look for further evidence linking ERVs to tumour development, with possible implications for therapies, and test the hypothesis that ERVs are PAMPs by seeing if individuals with higher levels of ERV expression exhibit a higher innate immune response. The work on ERVs in cancer involved the human ERV type-K HML2 lineage (HERV-K (HML2)), an ERV lineage found in humans, in Merlin-deficient tumours. These are schwannomas that arise from Schwann cells and for which effective drug therapy is urgently needed. The work on ERVs in inflammation involved the *Papio cynocephalus* ERV (PcEV), in rhesus macaques infected with simian immunodeficiency virus (SIV) infection.

The main outcomes are as follows: regarding HERV-K (HML2) in human schwannomas, (i) HERV-K (HML2) proteins are overexpressed in schwannoma compared to Schwann cells; (ii) these proteins are released from the tumour; (iii) regulation of HERV-K (HML2) expression in the tumour appears to involve the transcription factor TEAD; (iv) schwannomas are potentially treatable using anti-HERV-K (HML2) monoclonal antibodies and antiretroviral drugs since both decreased proliferation *in vitro*. Regarding PcEV in SIV-infected macaques: (i) PcEV is transcriptionally active; (ii) PcEV can be retrieved at low levels in the blood of some macaque animals; (iii) the levels of PcEV in cells correlates strongly with the strength

of the innate response as measured by cellular levels of STAT1 transcripts – an interferon-stimulated gene (ISG).

Other recent research has shown that human ERV lineages, namely HERV-W and HERV-H, have been co-opted and are involved in placentation and pluripotency during development, respectively. The present work suggests that ERVs are involved in a wide range of biological process and supports the need for further research into the biological significance of ERVs for their hosts.

2 TABLE OF CONTENTS

3 LIST OF ABBREVIATIONS

ADCC *: Antibody Dependent Cell-mediated Cytotoxicity*
AGS *: Aicardi-Goutières syndrome*
AKT *: Protein kinase B*
ALV *: Avian leukosis virus*
ART *: Antiretroviral therapy*
ATP *: Adenosine triphosphate*
BaEV *: Baboon endogenous retrovirus*
BDNF *: Brain-Derived Neurotrophic Factor*
BLV *: Bovine leukemia virus*
C/EBPalpha *: CCAAT/enhancer-binding protein alpha*
CA *: Capsid protein*
cAMP *: Cyclic adenosine monophosphate*
CAR-T cells *: Chimeric antigen receptor T cells*
cGAS *: Cyclic GMP-AMP synthase*
ChIP *: Chromatin immunoprecipitation*
COFAL *: Complement-fixation test for avian leukosis*
CREB : *cAMP response element binding protein*
CTF *: CCAAT box-binding transcription factor*
CTGF *: Connective tissue growth factor*
CTP *: Cytidine triphosphate*
DAMP *: Danger associated molecular patterns*
DNA *: Deoxyribonucleic acid*
dsDNA *: Double-stranded DNA*
EMSA *: Electrophoretic mobility shift assay*
Env : *Viral envelope glycoprotein*
ER *: Endoplasmic reticulum*
ErbB2 *: Epidermal growth factor receptor 2*
ERK *: Extracellular signal-regulated kinases*
FAK *: Focal adhesion kinase*
FeLV *: Feline leukemia virus*
GAPDH *: Glyceraldehyde-3-phosphate dehydrogenase, : Glyceraldehyde-3-Phosphate Dehydrogenase*
GMP *: Guanosine monophosphate*
GR : Glucocorticoid receptor
GTP *: Guanosine triphosphate*
HERV *: Human endogenous retroviruses*
HERV-K *: Human endogenous retrovirus type-K*
HERV-W *: Human endogenous retrovirus type-W*
hESC *: human embryonic stem cell*
HFV *: Human foamy virus*
HIV *: Human immnuodeficiency virus*
hNPCs *: human neural progenitor cells*
HSRV *: Human spumaretrovirus*
HTLV-1 *: Human T-cell lymphotropic virus type 1*
HYAL2 *: Hyaluronidase-2*
IAP *: Intracisternal A-type particles*
IFN *: Interferon*
IGF-1R *: Insulin-like growth factor receptor*
IL *: Interleukin*
IN *: Integrase*
ISG *: Interferon stimulated genes*
JSRV *: Jaagsiekte sheep retrovirus*
LATS 1/2 *: Large tumour suppressor kinases 1/2*
LINE *: Long interspersed nuclear element*
lncRNA *: Long non-coding RNA*
LTR *: Long terminal repeats*
MA *: Matrix protein*
MCS *: Multiple cloning site*
MMTV *: Mouse mammary tumour virus*
MOB1 *: MOB kinase activator 1*
MPMV *: Mason-Pfizer monkey virus*
mRNA *: Messenger RNA*
MS *: multiple sclerosis*
MST *: Macrophage stimulating protein*

MST1R *: Macrophage-stimulating protein receptor*
MuLV *: Murine leukimia virus*
mya *: million years ago*
NC *: Nucleocapsid*
NFI *: Nuclear Factor I*
NF-Y *: Nuclear transcription factor Y*
NGF *: Nerve Growth Factor*
NK *: Natural killer*
nt *: nucleotide*
OCT *: Octamer-binding transcription factor 4*
ORF *: Open reading frame*
p-AKT *: phosphor-AKT*
PAMP *: Pathogen associated molecular patterns*
PBMC *: Peripheral blood mononuclear cells*
PcEV *: Papio cynocephalus endogenous retrovirus*
p-ERK *: phospho-ERK*
p-FAK *: phosphor-FAK*
POU2F1 *: POU domain, class 2, transcription factor 1*
PR *: Protease*
RIG-1 *: Retinoic Acid-Induced Gene-1*
RNA *: Ribonucleic acid*
RNAi *: RNA interference*
RON *: Recepteur d'Origine Nantais*
RSV *: Rous sarcoma virus*
RT *: Reverse transcriptase*
RT-qPCR *: Reverse transcription polymerase chain reaction*
SA *: Splice acceptor site*
SAV1 *: Salvador homolog 1*
SD *: Splice donor site*
SDS-PAGE *: Sodium dodecyl sulphate–polyacrylamide gel electrophoresis*
SERV *: Simian endogenous retrovirus*
SLE *: Lupus erythematosus*
SOX2 *: Sex determining region Y-box 2*
SRV *: Simian retrovirus*
ssDNA *: Single-stranded DNA*
STAT1 *: Signal transducer and activator of transcription 1*
STING *: Stimulator of interferon genes*
SU *: Surface unit*
TAZ *: Transcriptional coactivator with PDZ-binding motif*
TEAD *: TEA domain family member*
TM *: Transmembrane unit*
TNF *: Tumour necrosis factor*
TNFα, *: Tumor necrosis factor alpha*
U *: Unique sequence*
UTP : *Uridine triphosphate*
VEGFR *: Vascular endothelial growth factor receptor)*
vlncRNA *: Very long non-coding RNA*
YAP *: Yes-associated protein*

4 LIST OF ILLUSTRATIONS AND TABLES

5 INTRODUCTION

5.1 Retroviridae

The *Retroviridae* appears as a family of viruses just after 1970, the date of the discovery of the reverse transcriptase (Baltimore, 1970; Temin and Mizutani, 1970). To briefly present the context, it was known that some oncoviruses (an old name for viruses causing cell transformation and cancer) such as the prototypic Rous Sarcoma Virus (RSV), were RNA viruses (Crawford and Crawford, 1961). In the 1960s it was discovered that virus production, by such virus-transformed cells, was inhibited using inhibitors of DNA synthesis, implicating an unknown DNA intermediate in the viral cycle (Bader, 1964). Howard Temin suggested that the genome of an RNA virus could be converted into DNA. Ultimately, an RNA-dependent DNA polymerase, termed Reverse Transcriptase, was independently discovered by Temin and David Baltimore, who both looked for and found the enzyme activity in purified virions. This finding went against the central dogma of molecular biology that genetic information always flowed from DNA to RNA to protein (Baltimore, 1970; Telesnitsky and Goff, 1997; Temin and Mizutani, 1970). In fact, during viral replication, a retrovirus reverse transcribes RNA into DNA. The DNA intermediate is then integrated in the genome and is defined as provirus. Another feature which defines retroviruses, along with other retroelements, is integration into the host DNA. It may even be their most important feature, as only retroviruses can become part of the host genome through infection.

5.1.1 Retrovirus viral particle

Members of the *Retroviridae* family share the following common features: looking from inside to outside, the virion is presented as two positive sense single-stranded RNA molecules, protected by a core composed of the capsid proteins, all surrounded by an envelope made of bilayer lipid membrane harboring viral envelope glycoproteins (Figure 1).

The viral RNA constitutes the viral genome. It is composed of an internal region coding for viral components (*gag-pro-pol-env*) flanked by 2 identical long terminal repeat (LTR) sequences: LTR-*gag-pro-pol-env*-LTR.

5.1.2 Genome of a member of the Retroviridae

The LTRs in the RNA stage of the viral life cycle are composed of R-U5 sequences at the 5' end, and U3-R at the 3' end. In the provirus, at each end, there is a U3-R-U5 sequence (Figure 2). U3 (Unique sequence 3) contains promoters such as the TATA box, and enhancers such as CAAT boxes along with responsive elements to transcription factors. Some responsive elements are shared among viruses, some are not. U3 also contains the polyadenylation signal (AATAAA). R (repeat sequence) usually contains transcription start and termination sites. U5 in some viruses possess responsive elements to transcription factors. The sequence that binds tRNA is used as a primer for the reverse transcription, and the sequence needed for genome packaging into the viral core is found between U5 and the first open reading frame (ORF) (Figure 2).

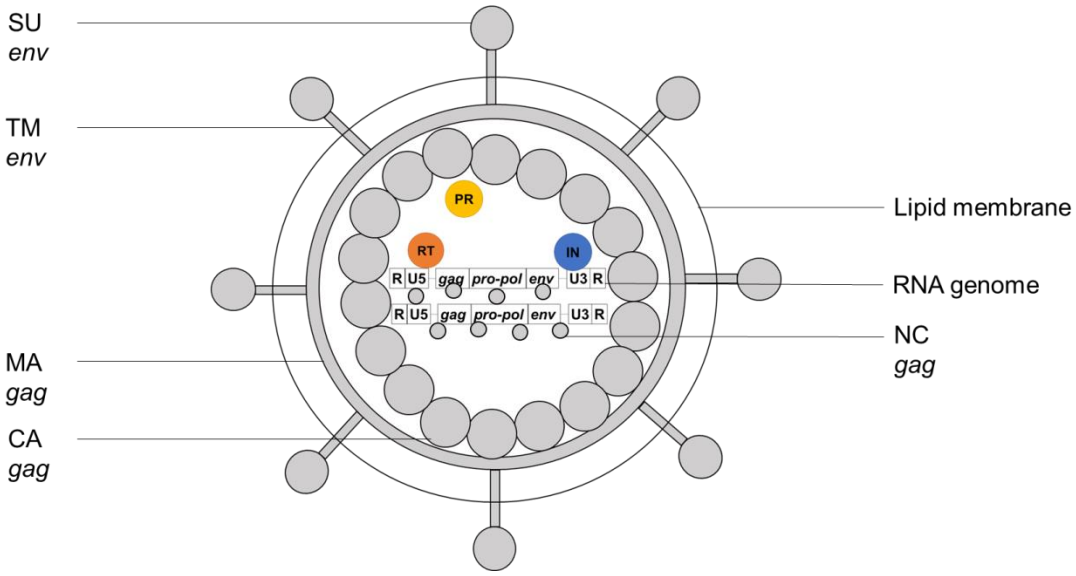


Figure 1: Schematic representation of a retroviral particle/virion.

MA: matrix; CA: capsid; NC: nucleocapsid; PR: protease; RT: reverse transcriptase; IN: integrase; SU: surface unit of the envelope glycoprotein; TM: transmembrane unit of the envelope glycoprotein. The virion is composed of two genomic RNAs linked to NC. Attached to the genomic RNAs, there are RT, tRNA and IN. PR cleaves Gag into CA proteins that assemble to constitute the viral core. The viral core is surrounded by a lipid membrane that harbors envelope glycoproteins, constituted by two subunits bound together, the SU and the TM. The core is linked to envelope by the MA.

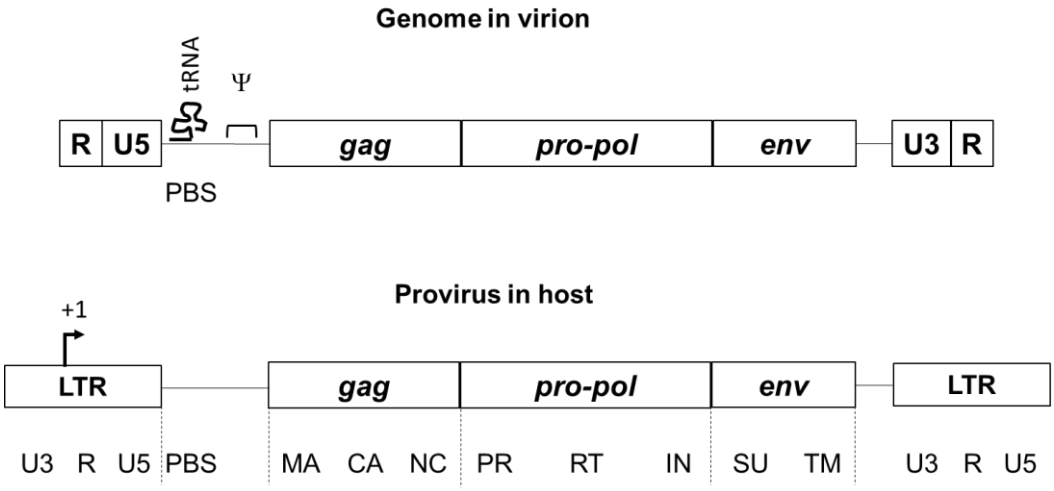


Figure 2: Structure of retroviral genome in virion and in the host.

In the virion, RNA genome is organized as follows: a short repeat sequence at each end named R; U5 is a unique sequence element after R at the 5' end; PBS is the primer binding site that bind the tRNA required to prime the reverse transcription, located between U5 and the beginning of the *gag* ORF; Ψ is the sequence required for RNA genome packaging into the viral core, then *gag*, *pro-pol* and *env* ORFs; and U3 which is a unique sequence element before R at the 3' end. In the host genome, the provirus has the same organization with complete Long Terminal Repeats (LTRs) that correspond to U3, R and U5. U3 usually contains the TATA box where transcription initiation complex bind, and a polyadenylation signal, which initiates transcription termination. At the 5' end, the start of R corresponds to the transcription start site (TSS), and the poly-A tail is added at the end of R at the 3' end.

The *gag* ORF generally encodes a polyprotein made of nucleocapsid (NC) which interacts with the genome, capsid protein (CA) which composed the core, and matrix protein (MA) which links the core to the envelope (Figure 1 and 2).

The *pro-pol* ORFs encode for the viral enzymes: the protease (PR) which cleaves Gag polyprotein into all sub-mature components described above (NC, CA, MA). The reverse transcriptase (RT) which converts RNA into DNA during the viral cycle. The integrase (IN) which integrates the provirus into the host genome (Figure 1 and 2).

The *env* ORF encodes for a glycoprotein present on the viral envelope (Env), which plays a role in viral entry. Env is composed by a surface unit (SU) which binds to surface receptors on host cells, linked to a transmembrane unit (TM) which mediates virion-host cell membrane fusion allowing the virus to enter into the cell (Figure 1, 2 and 3).

5.1.3 Viral cycle

The viral cycle can be presented in two phases: an early phase, which corresponds to events happening between the viral attachment to the cell and the integration of the provirus into the host genome; and a late phase during which the virus components are produced and packed to make new progeny viruses.

During the early phase, the SU binds to a cellular receptor present at the surface of the plasma membrane, this step is known as the viral attachment. Such an event produces a conformational change of the TM which reveals a fusion peptide responsible for joining the viral membrane and the plasma membrane of the host cell together resulting in a fusion of the two lipid membranes, allowing the viral core to be released inside the host cell. Following viral entry, the viral genome undergoes reverse transcription. The R-U5-*gag-pro-pol-env*-U3-R genomic RNA is converted to the U3-R-U5-*gag-pro-pol-env*-U3-R-U5 proviral DNA molecule. The Capsid with the newly formed DNA molecule migrates to the nuclear pore, where DNA is transported inside the nucleus. There, IN that is attached to the viral genome, enzymatically cuts the host genomic DNA and ligates the provirus in it. At the end of this phase, the provirus is part of the host cell genome.

The provirus functions as a host gene. LTRs contain host-like promoter sequences allowing the provirus to be transcribed as capped and poly-adenylated messenger RNAs (mRNA), which: on one hand constitute a pool of newly formed viral genomes, and on the other hand, provide a source of mRNAs for producing viral proteins. Two major mRNAs are distinguished: - an unspliced mRNA encoding both Gag and Gag-pro-pol polyproteins. Later in the viral cycle, during virion release from the cell, PR cleaves the Gag polyprotein into NC, CA, MA, and the Pol polyprotein into RT, IN (Figure 3, reviewed in Konvalinka et al. (2015)); - a spliced mRNA which spans *gag-pro-pol*, linking the *env* ORF to the 5'end of mRNA (Figure 3). The spliced mRNAs encode for the envelope glycoprotein Env. Env contains a signal peptide necessary for entry into the endoplasmic reticulum (ER) and subsequent protein glycosylation. The glycoprotein is likely further glycosylated in the Golgi apparatus, before being transported to the cell surface.

After the Gag polyprotein is produced, it assembles with the viral genome. The MA from Gag makes the link between the viral genome-Gag polyprotein complex and the envelope glycoproteins. Finally, the new viruses bud from the plasma membrane which provides a lipid bilayer membrane to the virions. After budding, the protease cleaves Gag into mature NC, CA, and MA to raise mature and infectious retroviral particles. Inhibiting the protease results in non-infectious particles (Kohl et al., 1988).

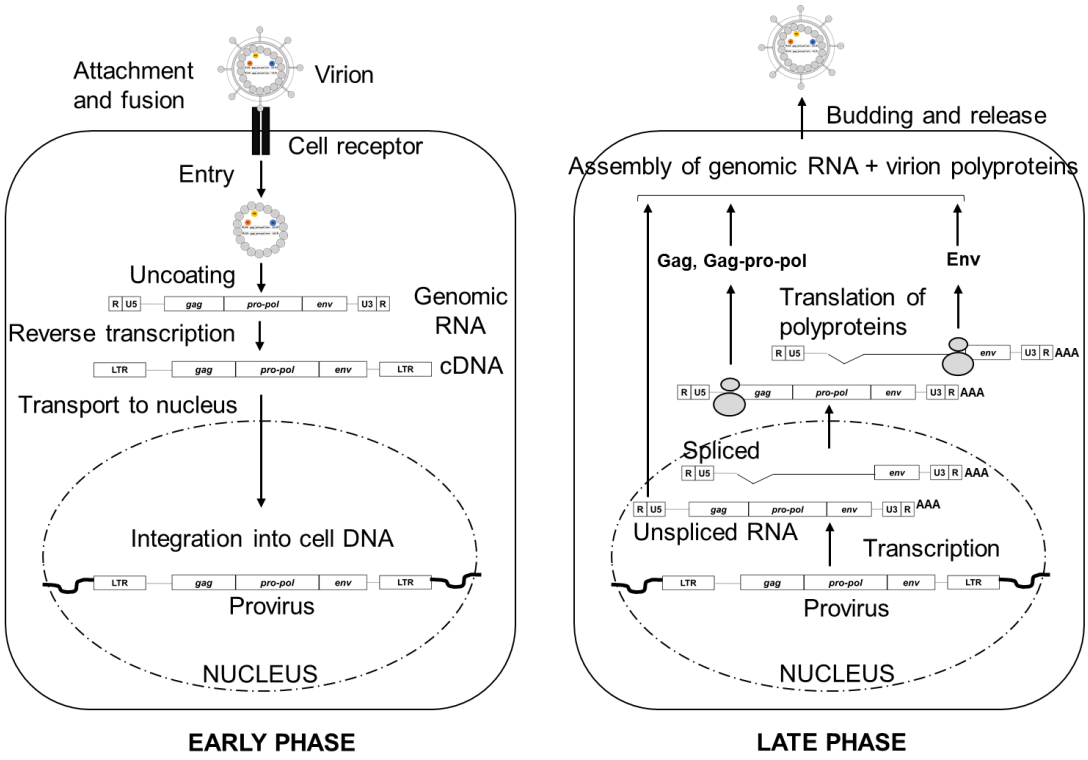


Figure 3 : Replication cycle of a prototype retrovirus

The viral particle attaches to the cell through interaction between the SU and a cellular receptor. Such interaction is followed by a fusion of the virus and the cell membrane, and the entry of the core in the cytoplasm. The core is uncoated and the RT reverse transcribes the genomic RNA into DNA. The viral DNA is transported in the nucleus and integrated to the host genome by IN. The provirus contains transcription motifs on its LTRs, able to recruit cellular transcription machinery. LTR-driven transcription produces unspliced RNAs that serve as genomic RNA for new virions, and are used to produce Gag, Pro and Pol polyproteins; as well as a spliced RNAs that are translated into the Env. Gag and Gag-pro-pol polyproteins are synthesized by free cytoplasmic polysomes, they assemble with the genomic RNA in the cytoplasm to form an immature core structure. Env is synthesized in the ER and is modified by glycosylation and cleavage (to give the SU and the TM, joined together) by cellular furin proteases, it is then inserted into the cell membrane. The immature core is linked to Env glycoproteins by the MA. The core buds and is engulfed by a piece of the cell membrane. During the process, PR cleaves Gag and Gag-pro-pol polyproteins into MA, CA, NC, RT, and IN

to form mature virion. Inactivating the PR usually results in non-infectious virions. Adapted from *The Retroviridae*, p.160 (Luciw and Leung, 1992).

5.2 Classification

Over time, there have been three main ways of classifying retroviruses, each bearing witness to historical decisions within the virology community. Such changes in classification were due to the understanding that two viruses could cause a close-related disease, but, at the same time, be very distinct – e.g. Both Avian leukosis virus (ALV) and Human T-cell lymphotropic virus type 1 (HTLV-1) cause blood malignancies in their respective hosts, but, in respect of the updated classification, ALV is a simple retrovirus belonging to the genus of *Alpharetroviruses*, while HTLV-1 is a complex retrovirus belonging to the *Deltaretroviruses*. Furthermore, it appears that the focus should be on the structure of the genome to define new and closely related viruses, rather than the clinical outcome.

5.2.1 Classification based on virion morphology

In total, six distinct morphologies of viral particles have been described: A-type, B-type, C-type, D-type, lentiviral type, and spumaviral type particles (figure 4) (Luciw and Leung, 1992). Although, in the past, retroviruses used to be distinguished by the type of particles they produced, do note that this classification is still used in parallel sometimes, as morphological structure of retroviruses using electron microscopy provides additional information for describing the virus.

A-type

A-type particles are distinguishable from all the others as they are strictly intracellular. They resemble a condensed circular structure in cell cytoplasm and sometimes within the ER, which can be explained by the fact that such retroviruses lack the *env* ORF. Members of this type are usually endogenous – e.g. mouse Intracisternal A-type Particles (IAP).

B-type

B-type particles are present as an enveloped particle with a condensed hexagonal core, which is located at one side of the particle (eccentric). The intracellular phase resembles

that of A-type particles. Budding of B-type particles can occur once the core is fully assembled, something which occurs intracellularly. Both endogenous and exogenous viruses can produced B-type particles – e.g. Mouse mammary tumour virus (MMTV, both endogenous and exogenous).

C-type

C-type particles are produced as enveloped particles with a central, electron-dense core. They differ from A-type, B-type and D-type as the core appears to assemble during the budding process rather than beforehand. These viruses are both exogenous and endogenous. – e.g. HTLV-1 (exogenous), Murine leukemia Virus (MLV/MuLV) (endogenous and exogenous).

D-type

D-type particles are enveloped particles with a 'rod-shaped' core. The core is assembled before the budding event. These viruses are usually exogenous. – e.g. Mason-Pfizer monkey virus (MPMV); Simian retrovirus (SRV)

Lentivirus

Lentiviral particles resemble D-type particles. However, the shape of the core in the virion is more conical in structure, and the core assembles while the virus buds. The prototype virus here is the Human Immnunodeficiency Virus (HIV). *Lentiviruses* are exogenous only.

Spumavirus

Spumaviral particles resemble the structure of C-type particles. The core is central in the virion, but it assembles before the virus buds. Another difference is the virion membrane which appears denser, suggesting a higher glycosylation state of the envelope glycoproteins. The prototype virus is Human foamy virus/Human spumaretrovirus (HFV/HSRV).

5.2.2 Old classification

The old classification was based on the outcome of viral infection on the host, hence three main subfamilies were distinguished: the *oncovirinae* as causing malignancies,

the *lentivirinae* as causing chronic and long-lasting infections resulting in cytopathic diseases (‘lenti’ is latin for slow), and the *spumavirinae*, for which there is not a known disease condition (Coffin, 1992).

5.2.3 New classification

The new taxonomy builds links between viruses with similar genomic traits: organization of *gag-pro-pol* reading frames, presence of accessory proteins (that take part to the infection, but are not packed in the particle. Also, their existence sets the boundary between simple and complex retroviruses: simple retroviruses are devoid of accessory proteins, complex retroviruses possess some), and sequence similarity. Such classification is based mainly on building phylogenic trees from sequences of retroviruses. The new classification raised up seven genera of retroviruses: *Alpharetroviruses* – Avian leukosis virus (ALV)-related such as RSV; *Betaretroviruses* – MMTV-related as Human endogenous retrovirus type-K (HERV-K), Jaagsiekte sheep retrovirus (JSRV); *Gammaretroviruses* – MLV-related, Feline leukemia virus (FeLV), Human endogenous retrovirus type-W (HERV-W), Baboon endogenous retrovirus (BaEV), Simian endogenous retrovirus (SERV); *Deltaretroviruses* – Human T cell lymphotropic virus (HTLV), Bovine leukemia virus (BLV); *Epsilonretroviruses* – Walleye dermal sarcoma virus (WDSV), Walleye epidermal hyperplasia virus (WEHV); *Lentiviruses* – Human immunodeficiency virus (HIV), Simian immunodeficiency virus (SIV), Equine infectious anemia virus (EIAV), Feline immunodeficiency virus (FIV), Bovine immunodeficiency virus (BIV); *Spumaviruses* – Human foamy virus (HFV), also known as Human spumaretrovirus (HSRV), Bovine foamy virus (BFV), Simian foamy virus (SFV).

Today, the *Retroviridae* family is classified under the order of *Ortervirales* which harbor all reverse-transcribing viruses except *Hepadnaviridae* (Hepatitis B viruses) (Krupovic et al., 2018). The *Retroviridae* include two subfamilies: (i) the *Orthoretrovirinae* which possesses six genera – *Alpharetrovirus*, *Betaretrovirus*, *Gammaretrovirus*, *Deltaretrovirus*, *Epsilonretrovirus* and *Lentivirus*; and (ii) the *Spumaretrovirinae* which possesses five genera – *Bovispumavirus*, *Equispumavirus*, *Felispumavirus*, *Prosimiispumavirus* and *Simiispumavirus* (International Committee on Taxonomy of Viruses (ICTV)).

Table 1 : Comparison between "old" and "new" taxonomy.

Adapted from *the Retroviridae*, p.43 (Coffin, 1992).

Taxon	Old taxonomy	New taxonomy
Family	Retroviridae	Retroviridae
Sub-family	*Oncovirinae*	*Orthoretrovirinae*
Genus		*Alpharetrovirus (ALV-related)*
e.g. species	ALV, RSV	ALV, RSV
Genus		*Betaretrovirus (MMTV-related)*
e.g. species	MMTV, JSRV	MMTV, JSRV
Genus		*Gammaretrovirus (MLV-related)*
e.g. species	MLV(MuLV), FeLV, HERV-W	MLV(MuLV), FeLV, HERV-W
Genus		*Deltaretrovirus (HTLV-BLV group)*
e.g. species	HTLV1,-2, BLV	HTLV1,-2, BLV
Genus		*Epsilonretrovirus*
e.g. species		WEHV, WDSV
Sub-family	*Lentivirinae*	
Genus		*Lentivirus*
e.g. species	HIV, SIV, EIAV, FIV, BIV	HIV, SIV, EIAV, FIV, BIV
Sub-family	*Spumavirinae*	*Spumavirinae*
Genus		*Spumavirus*
e.g. species	HSRV(HFV), BFV, SFV	HSRV(HFV), BFV, SFV

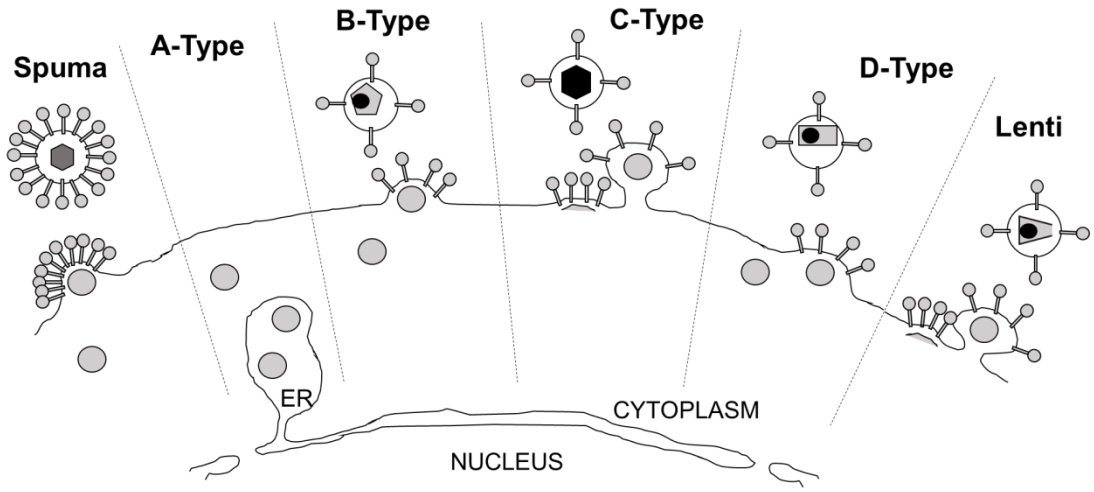


Figure 4: Virion morphologies.

Schematic representation of retroviral morphologies based on electron microscopy of mature virions and assembly intermediates in infected cells. Gray circle in cytoplasm or budding membranes depicted immature core. Viral-coded envelope glycoproteins are inserted in the membrane and appears as knobs on the external surface of cells where virions bud. A-type particles, are exclusively intracellular, found inside ER cisternae or free in the cytoplasm. B-type particles, show intracellular assembly of the core before budding, and mature virions possess an eccentric core that is tightly condensed. The intracellular form of C-type particles is an electron-dense crescent juxtaposed next to the cell membrane, and the mature virion presents a centrally located concentric core. D-type particles, show intracellular assembly of the core before budding, and mature virions possess an eccentric core in form of a cylinder with an electron-dense mass inside. The intracellular form of lentiviruses is an electron-dense crescent juxtaposed next to the cell membrane like C-type particles, and the mature virions present a core in shape of a truncated cone with an electron-dense mass inside. Spumaviruses show intracellular assembly of the core before budding, and mature virions possess a centrally located small concentric core, smaller in comparison to C-type, and highly prominent knobs. Adapted from *The Retroviridae*, p.170 (Luciw and Leung, 1992).

5.3 The discovery and classification of HERVs

5.3.1 The discovery of HERVs

The knowledge of Rous Sarcoma Virus (RSV), an RNA-based, cancer-causing virus in chickens was established in 1961 (Crawford and Crawford, 1961). Chickens were originally diagnosed for viral infection through the presence of viral antigens by the complement-fixation test for avian leukosis (COFAL) test. However, it was observed that some apparently healthy chickens were positive in these antigen detection tests. Payne and Chubb (1968) showed that the viral antigens in these chickens were possibly encoded by the genome. The authors showed inheritance of a viral antigen controlled by a single autosomal dominant gene, as chickens crossbred from an antigen-positive and -negative pure lines were positive. During the same period, it was common to culture chicken fibroblasts transformed with RSV. It is interesting to note the fact that the viral strain of RSV with that transforming capacity was the BRYAN-strain (Weiss, 2006). This strain does not harbor a gene encoding the viral Env. It is replaced by a sequence encoding the Src protein, responsible for oncogenic cell transformation (Jong et al., 1992). Cells infected with the BRYAN-strain required co-infection with a helper virus, the Avian Leukosis virus (ALV), to release RSV viral particles capable of reinfection (Hidesaburo Hanafusa, Teruko Hanafusa, 1963). In theory fibroblasts transformed with RSV cannot release newly-formed viral particles without ALV co-infection. However, viral particle release from transformed cells in culture, free of helper virus was described (Weiss, 1967) and ALV particles are indistinguishable from RSV when observed under an electron microscope (Dougherty and Di Stefano, 1965). Here came the second evidence of potentially endogenous retroviral sequence. The authors suggested that the Env necessary to release RSV particles could have an endogenous origin. This cryptic RSV released in culture free of helper virus was named RAV-0 (Vogt, 1967). It was in 1979-80, that the evidence proving the existence of retroviral-like sequences, responsible for the viral particles produced in healthy chickens, was finally produced by Astrin and colleagues (Astrin et al., 1980, 1979; Astrin and Robinson, 1979). They harvested DNA from many chickens, separated it using electrophoresis, and transferred DNA to nitrocellulose membrane for hybridization. They designed a probe from RSV sequence and detected endogenous viral loci (termed *ev*) in the chickens' DNA. More interestingly, the chickens which harbor one specific *ev*, designated *ev* 2, produced viral particles. By this, the authors

conclude *ev* 2 codes for RAV-0. The discovery of endogenous retroviruses emerged from these virology and genetics experiments in chickens.

Around the same time, similar findings were obtained in mice. Spontaneously-released viral particles were described in cultures of murine cells. In 1969 Aaronson et al. showed that some cell lines derived from BALB/c mouse embryos were able to release murine leukemia virus (MLV) (Aaronson et al., 1969). MLV is a retrovirus first discovered in 1951, associated with leukemias and lymphomas (Gross, 1951). The fact that MLV is spontaneously released in culture suggested MLV to be encoded by the genome, hence the genetic analysis of viral gene expression of endogenous MLV was studied in great details in the 1970s and 1980s (Weiss, 2006). Another retrovirus, the Mouse mammary tumour virus (MMTV), associated with breast cancer in mice, was also discovered to be endogenous. It was reported that several mice of the O20 strain, free of exogenous MMTV, treated with urethane (highly carcinogenic) and irradiated, contracted mammary tumours within a year; the tumours presented retroviral-like particles observed using electron microscopy (Timmermans et al., 1969). These findings highly suggested that the genome of the O20 mouse encoded a retrovirus, of which the production could be induced. The retrovirus released spontaneously was likely the MMTV, since tumour cell-free extracts from treated-mice and injected in BALB/c mice, induced mammary tumour development in the latter (Timmermans et al., 1969).

Endogenous retroviruses have many times been linked with tumours. Taking that into account, humans were screened for endogenous retrovirus. The first evidence of human endogenous retroviruses (HERVs) was established by Callahan et al. (1982) in 1982 and Ono (1986) in 1986. They detected HERV in humans by hybridization of human DNA using probes derived from MMTV and Hamster Intracisternal A particle (IAP), respectively. IAP is an A-type endogenous retrovirus present in rodents such as mice and Syrian hamsters.

5.3.2 The naming and classification of ERVs

In general, ERVs have been named after the species in which they were discovered, by adding one or two letters referring to the species before "ERV" (Stoye and Coffin, 1987). Examples are Human ERV (HERV), Mouse ERV (MERV or MuERV), and Rabbit ERV (RERV). An issue exists related to the fact that one name may refer to different species ERV, an example is the CERV that can refer to Chimpanzee ERV (Polavarapu et al., 2006) or Crocodilian ERV (Jaratlerdsiri et al., 2009). To over-come this problem, consideration are to be made towards the use of the scientific for each species, and a new nomenclature has been proposed (Gifford et al., 2018).

Regarding the classification, it has been driven by the methodology and wet-lab techniques available over the last 30 years from their discovery. Initially, since they have been discovered using low stringent hybridization, ERVs were grouped with the virus they were similar too (reviewed in Escalera-Zamudio and Greenwood (2016); Gifford et al. (2018)). For example, MMTV-based probe can bind some HERV sequences (Callahan et al., 1982) and are part of the same "family". In fact, "HML" stands for human MMTV-like.

Next, a classification based on the complementary tRNA that serves as primer for the reverse transcription was attempted. Incidentally, HERV-K was named after the tRNA that correspond to the lysine. Today, this system of classification is obsolete and is rather unsuitable as it is clear now that retroviral members of the same "lineage" (e.g. HERV-Ws) does not necessarily use the same tRNA to prime the reverse transcription (Grandi et al., 2016).

Within the genomic era which allows the sequencing of the genome and *in silico* sequence analysis, it is possible to realize phylogeny and evolution studies of ERVs, making for example a the relationship within a lineage more clear. In fact, by comparing sequences for retroviral genes it is possible to imply a parental relationship (link several retroviruses to a common ancestor) and predict all genomic proviruses that have possibly arisen from the same retroviral infection event in an ancestor (reviewed in Gifford et al. (2018)). There has been an attempt to classify ERVs as it is done for exogenous retroviruses: thus, all ERVs that are related to the *Gammaretrovirus* are

grouped in the “ERV class I” family; the ones related to the *Betaretrovirus* are grouped in the “ERV class II” family; and the ones related to the *Spumaretrovirus* are grouped in the “ERV class III” family (figure 5) (reviewed in Gifford et al. (2018)).

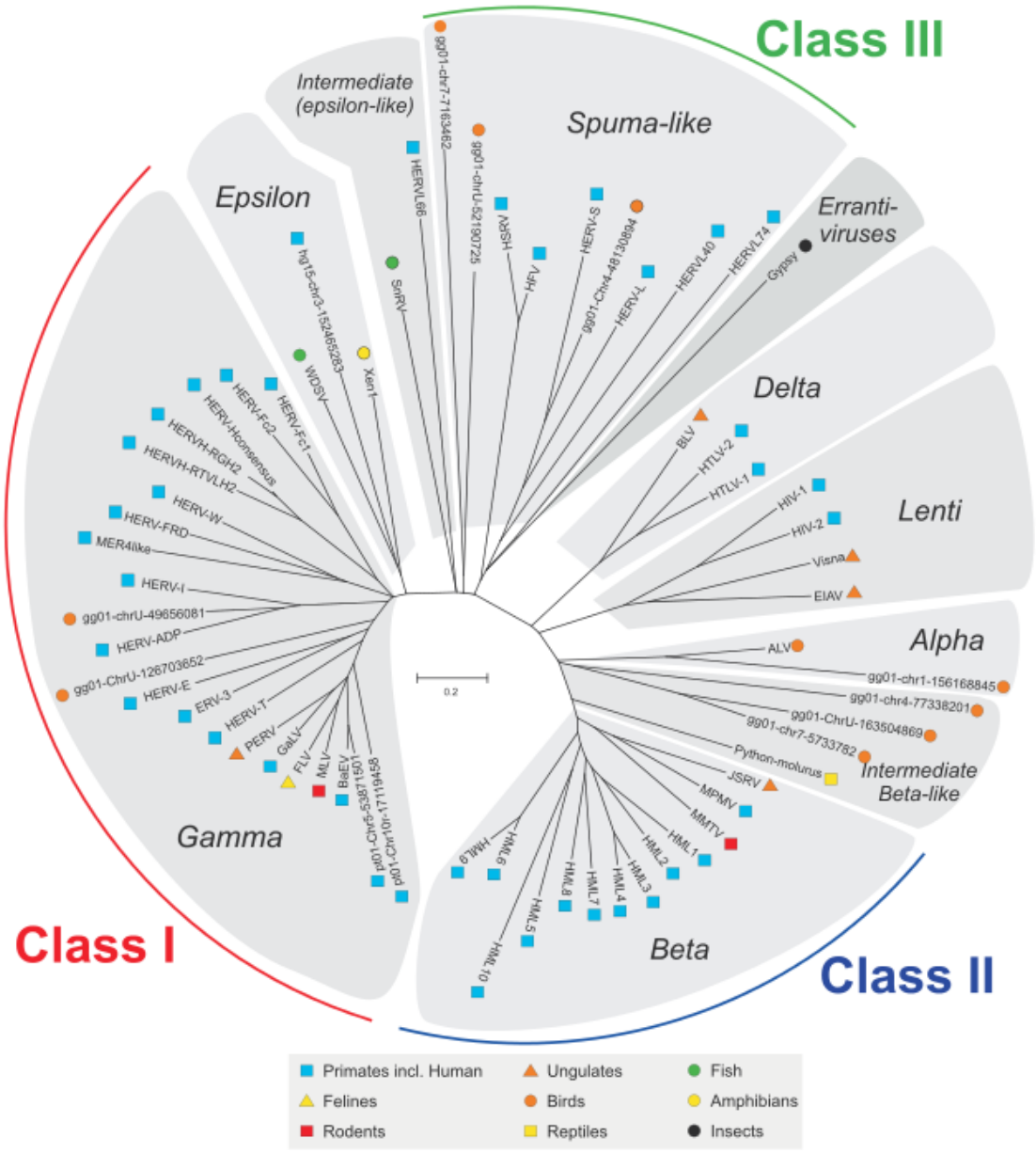


Figure 5: ERV classification

The main groups of ERVs are class I, II and III which are related to the Gammaretrovirus, Betaretrovirus and Spumaretrovirus, respectively. The figure is from Jern et al., (2005).

5.4 HERV-K (HML2) genome

A total of 91 HERV-K (HML2) proviruses can be found in the human genome and are classified in two types: type 1 that represents 26% of all proviruses and possess a 290-nucleotide (nt) deletion at the junction between *pol* and *env* ORF (nt 6501-6790); type 2 represents 74% of all proviruses and lacks the deletion (Ono et al., 1986; Subramanian et al., 2011).

A prototypic HERV-K (HML2) viral genome is 9469nt long and consists of a LTR sequence (nt 1-968); followed by an untranslated region from nt 969-1111 that contains the PBS (nt 971-988); *gag-pro-pol* ORFs (nt 1112-6701), with pro and pol being translated via two ribosomal frameshifts (Figure 6) ; *env* ORF (nt 6701-8790), and another LTR sequence (nt 8502-9469) (Ono et al., 1986; Turner et al., 2001). The LTRs possess promoter capacity, able to initiate transcription at the 5' end of the viral genome. Multiple responsive elements were reported, such as binding sites for MITF-M, NF-AT, NF-κB, Oct4, Sp1 (GC boxes) and YY1 (Fuchs et al., 2011; Gonzalez-Hernandez et al., 2012; Grow et al., 2015; Katoh et al., 2011; Knössl et al., 1999). The transcription start site was reported at position nt 793; and the transcription termination site at position nt 878, suggesting a length of 792nt for U3 (nt 1-792 at 5'; nt 8502-9294 at 3'), 84nt for R region (nt 793-876 at 5'; nt 9295-9378 at 3'), 92nt for U5 (nt 877-968 at 5'; nt 9379-9469 at 3') (Fuchs et al., 2011).

Altogether, HERV-K (HML2) proviruses produce four main transcripts (Figure 6):
(a) An unspliced transcript which encodes for the polyproteins Gag and Gag-pro-pol. In Type 2, that transcript possesses a stop codon after *pol* and does not translate the *env*. In Type 1, it is ambiguous but there is a suggestion that it could produce a pol-env polyprotein (Chen et al., 2013).

(b) A spliced transcript, which results from '(a)' after splicing of the region corresponding to *gag-pro-pol*. The splice donor site 1 (SD1) and splice acceptor site 1 (SA1) are at positions nt 1077 and 6433, respectively (Mayer et al., 2004). Type 1 possess the 292nt deletion about 7nt after SA1, so it lacks the starting ATG of *env* located at 6451 and may not produce the full length Env glycoprotein through the spliced transcript '(b)'(Ono et al., 1986).

(c) A double-spliced transcript that encodes for an accessory protein named Rec. It is responsible for transcript export from the nucleus to the cytoplasm, in a similar fashion as HIV Rev and HTLV Rex proteins (Magin-Lachmann et al., 2001; Magin et al., 1999). This transcript can be built from Type 2 proviruses only. It results from a second splicing of '(b)' at positions nt 6711 and 8411, for the second SD/SA pair (SD2 and SA2) (Mayer et al., 2004).

(d) A double-spliced transcript raised from Type 1 proviruses only encodes Np9, defined as a HERV-K (HML2) oncogenic protein (Chen et al., 2013). It is produced from '(b)'by a second splicing event that involves a SD (SD3) and SA at positions nt 6494 and 8117, respectively (Armbruester et al., 2002). Note that the SA at 8117 in type 1 HERV-K (HML2) proviruses corresponds to SA2 at 8411 (nt 8411) on type 2 proviruses.

Gag ORF encodes a 75kDa polyprotein, precursor of MA (15.3kDa), SP1 (1.5kDa), p15 (15kDa), CA (27.7kDa), NC (10kDa), QP1 (2.5kDa) and QP2 (2.1kDa), that shapes the core of the retroviral particle (George et al., 2011).

Pro encodes the Protease (nt 2912-3914, which is equivalent to a 334 amino-acid length sequence, when separately considered from Gag) that processes Gag polyprotein.

Pol encodes the viral enzymes that possess the RT activity (nt 3998-4624, which is equivalent to a 209 amino-acid sequence) responsible for the reverse transcription of the viral RNA into DNA upon viral entry into the host cell; and the integrase activity (nt 5630-6169, which is equivalent to a 180 amino-acid length sequence) that integrate the DNA synthetized by RT into the host genome (Ono et al., 1986).

Env encodes a 70kDa polyprotein that consists of a signal peptide (SP) reported to be 13kDa in size (Ruggieri et al., 2009), SU (44kDa) and TM (26kDa) domains. The signal peptide directs the Env polyprotein to the ER, where the latter undergoes glycosylation. Approximately 11 glycosylation sites can be reported using the rule N-X-S/T (Asparagine-any amino acid-Serine/Threonine) with X different of a P (Proline) (as reported in UniProtKB, Q69384, corresponding to HERV-K108 *env*, termed also ERVK6). Considering glycosylation, the size of Env range from 70-95kDa, with SU

ranging from 42-55kDa and TM from 26-38kDa (Hanke et al., 2009). A group also reported an over-glycosylated TM with a size of 55kDa (Kämmerer et al., 2011).

Rec encodes a 14.5kDa protein and Np9 is approximately 9kDa.

5.4.1 HERV-K (HML2) viral cycle/replication

A retroviral cycle is described as follows: (a) attachment of envelope glycoproteins and viral entry in the host cell, (b) reverse transcription of the viral RNA genome by RT, (c) integration in host genome, (d) transcription driven by the LTRs and translation of new viral proteins, (e) assembly of the viral proteins, (f) budding and release of new virions from the host cell.

HERV-K (HML2) has a very weak replication capability, defined here as the ability to produce viral particles able to infect and integrate the host genome and then repeat the process. Firstly, only one *env* sequence coming from HERV-K108 was demonstrated to be functional, meaning it can attach and allow cell-virus fusion (Dewannieux et al., 2005). Secondly, few HERV-K (HML2) proviruses appear to encode a functional RT (HERV-K113; HERV-K10). Also, the RT activity was more than three times lower than that of the MLV and AMV (avian myeloblastosis virus), replication-competent retroviruses used as positive control in the studies (Berkhout et al., 1999; Contreras-Galindo et al., 2017). Thirdly, even though a functional integrase encoded by HERV-K10 (Kitamura et al., 1996) was described, a recent study suggest that HERV-K (HML2) viruses can be transmitted but do not integrate in the genome of newly infected-host cells (Contreras-Galindo et al., 2015). Finally, although there is transcription from LTRs (Hanke et al., 2016; Hohn et al., 2013), HERV-K (HML2) proteins could mainly be detected in human tumours, stressing that HERV-K (HML2) proteins can assemble and produce particles only in specific conditions. In fact, HERV-K (HML2) virions were described in melanoma, breast cancer and lymphoma (Hohn et al., 2013). Also, there is no HERV-K (HML2) viremia, as no particles can be found in the blood of human donors (Bhardwaj et al., 2014). These findings correlate with the relatively low number of HERV-K (HML2) insertions in the modern human genome (Marchi et al., 2014). Despite the fact that viral-like particles have been observed in

specific conditions (Boller et al., 2008), HERV-K (HML2) is considered to be "extinct" as they have likely stopped replicating (Magiorkinis et al., 2015).

Nonetheless, despite having many defective loci, HERV-K (HML2) is unusual for a human ERV in that it is the only one that is less degraded and still possesses a full-length ORF for all genes (taking all loci together). This is not the case in other species such as the mice which have lots of recently integrated elements in them (illustrated in Fig S4).

5.5 PcEV genome

The following description is based on the work by Mang et al. (1999) as they discovered the *Papio cynocephalus* endogenous retrovirus (PcEV) from kidney tissue of a yellow baboon (Mang et al., 1999); and complementary information gathered from NCBI database using PcEV accession number AF142988. An estimate of ~38 to 50 proviruses were found in the genome from the baboon from which it was discovered. PcEV belongs to the *Gammaretroviruses*, its viral genome is 8572nt long and consists of a LTR sequence (nt 1-510); followed by an untranslated region from nt 510-962 that contains the PBS (nt 513-530); *gag-pro-pol* ORFs (nt 963-6165), with *pro* and *pol* being translated via a suppression of termination – i.e. the ribosome readthrough the stop codon – (Figure 6) (Yoshinaka et al., 1985); *env* ORF (nt 6111-8057), and another LTR sequence (nt 8084-8572). The LTRs possess promoter capacity and can initiate transcription at the 5' end of the viral genome. On the 5' LTR, four enhancer sequences of 21nt, termed direct repeats (DR) were reported (nt 45-65; nt 66-86; nt 87-107; nt 108-123), with the fourth one being incomplete (16nt instead of 21). The 3' LTR lacks the second DR (due to a deletion) and is therefore 21nt shorter. Such enhancers harbor binding sites for the transcription factor GATA-1. The LTRs also displayed four CCAAT boxes starting at positions nt 124, 161, 242, 281. CCAAT are potent enhancers for retroviral transcription (Lee, 2003). Some GC stretches can be found too suggesting potential binding sites for Sp1. The TATA box was found at position nt 351 and the polyadenylation signal AATAAA is at nt 8385. U3 was estimated from nt 1-377 and 8084-8440; R from 378-443 and 8441-8506; and U5 from 444-510 and 8507-8572.

As in other *Gammaretroviruses*, PcEV proviruses should produce two main transcripts (Figure 6B): (a) An unspliced transcript which encodes for the polyproteins Gag and Gag-pro-pol. (b) A spliced transcript, which results from '(a)' after splicing of the region corresponding to *gag-pro-pol*, the splice donor site (SD) is at positions nt 619.

Gag ORF (nt 863-2591) encodes a 60kDa polyprotein, precursor of MA, p12, CA, NC, that shape the core of the retroviral particle.

Pro ORF encodes the protease (nt 2570-2955, which is equivalent to a 128 amino acid sequence, when separately considered from Gag) that process Gag polyprotein.

Pol ORF encodes the viral enzymes that possess the RT and endonuclease activities (nt 2956-6165).

Env ORF (nt 6111-8057) encodes a 648-amino acid glycoprotein of approximately 70kDa. It consists of an SU (479 amino acids, ~50 kDa) and TM (166 amino acids, ~20 kDa). The SU harbor five potential glycosylation sites at amino acid residue 60, 310, 342, 345 and 381. The TM possesses a putative immunosuppressive peptide of 26 amino acids (residues 521-546).

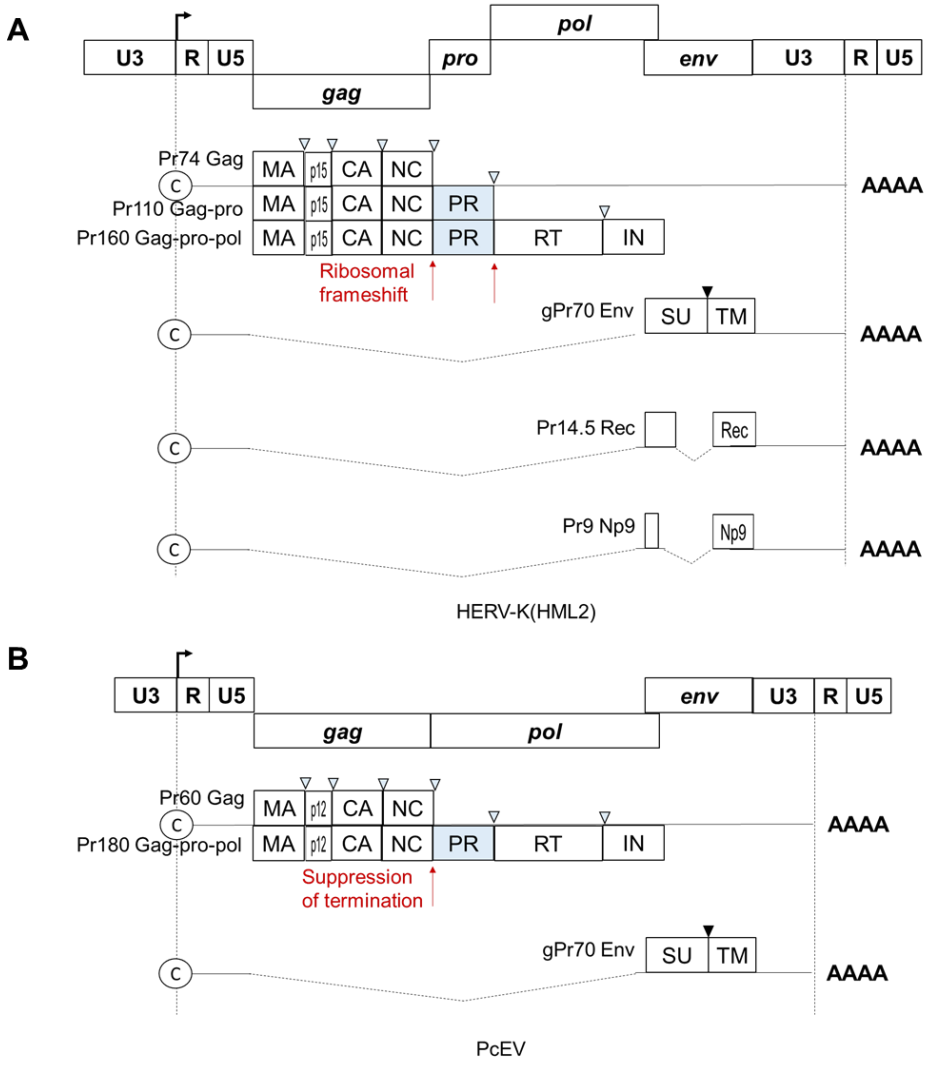


Figure 6 : Schematic representation of ORFs and transcripts from HERV-K (HML2) and PcEV

(A) HERV-K (HML2) transcribes an unspliced RNA that harbor *gag-pro-pol* ORFs, with pro and pol enzymes being translated by ribosomal frameshifts at the end of *gag* and *pro*, respectively. Env is translated from a spliced transcript. HERV-K (HML2) also produce double-spliced RNAs encoding for Rec in type 2 loci, and Np9 in type 1 loci (harboring 292nt deletion in *env*). (B) PcEV transcribes an unspliced RNA that harbor Gag-pro-pol ORFs, with pro and pol enzymes being translated by suppression of a stop codon, ribosome read through the stop codon and produce a Gag-pro-pol polyprotein that is cleaved into Gag, pro and pol enzymes by viral PR. Env is produced through a spliced RNA. For (A) and (B), Env subunit are separated by cellular proteases, cleavage site is indicated by a black arrow (▼). PR cleavage sites are indicated as blue arrows (▽).

5.6 HERV expression

The expression of HERVs relies on the capacity of their LTR sequence to retain a viable promoter activity, such as possessing binding sites for RNA polymerase and transcription factors (TATA box and enhancers), as well as not being subject to methylation events (histone and DNA methylation).

5.6.1 Methylation

Methylations of DNA and a certain lysine (K) on H3 histone (H3K9 and H3K27) are mechanisms of gene silencing. Such mechanisms are responsible for the silencing of many ERV loci. It happens on an evolutionary time scale (inactivation of ERVs occurring throughout history) as the state of silencing generally increases with the amount of time ERVs have spent in the genome since infection and integration. In fact, retroviral sequences from HERV-L are older than those of HERV-K, and H3K9me3, that is H3 methylated on the 9^{th} lysine, were showed to be associated with more sequences from HERV-L than HERV-K (Brattås et al., 2017). Methylations also happen within lifetime of individuals as each stage of the development is more or less permissive to the expression of ERVs (Fasching et al., 2015). Interestingly, the methylation of ERVs has been shown to be reversible, using cell reprogramming (Göke et al., 2015).

The family of proteins that have been suggested to be involved in the methylation of ERVs, are the Krüppel-associated box domain zinc finger proteins (KRAB-ZFP). Throughout evolution, the genes encoding these proteins appeared in mammals in parallel to ERV integration events. It is likely that, in the past, infection by retroviruses that can become endogenized would have been harmful for the offspring. Silencing of ERVs would have been critical for surviving, as suggested by the acquisition, alongside integration events, of those transrepressor proteins (Hurst and Magiorkinis, 2017).

Recently, KRAB-ZFP transrepressor activity was suggested to involve protein complexes with KAP1, also known as TRIM28. In fact, KAP1 interacts with KRAB-ZFP and the latter binds to DNA (Hurst and Magiorkinis, 2017). KAP1 knock-out resulted in ERV reactivation and transcription in mouse and human neural progenitor cells (hNPCs) (Fasching et al., 2015; Rowe et al., 2010). Such complexes are believed

to bind directly to or close to the LTRs and mediate methylation of H3. In fact, chromatin-immunoprecipitation experiments that pull down a methylated form of H3 (H3K9me3) revealed enrichments in sequences from HERV lineages (Brattås et al., 2017).

Regarding HERV-K (HML-2), it is highly likely that DNA methylation of LTRs and methylation of the lysines of H3 repress transcription. Cloning of LTR promoters from loci that are transcriptionally active *in vitro*, followed by bisulfide sequencing (CpG islands sequencing), revealed low levels of methylation (0-36%), while untranscribed loci *in vitro* showed higher levels of methylation (59-71%) (Lavie et al., 2005). Also, in the long term, the methylated promoters are suggested to undergo mutations which likely render it permanently inactive (Lavie et al., 2005). In addition, some sequences from HERV-K are enriched in ChIP experiments that pull down H3K9me3 in hNPCs (Brattås et al., 2017).

HERVs were found to be associated with specific developmental stages during embryogenesis: LTR14B for 2-cell stage, LTR12c for 4-cell, HERV-L for 8-cell, HERV-K (HML2) for morula and HERV-H for blastocyst (Göke et al., 2015). HERV-K (HML2) maintain higher levels of transcription in comparison to the others lineages in tissues from adults, even though the transcription at that stage is limited, as revealed by RNAseq analysis of tissues from adults (adipose, adrenal, brain, breast, colon, heart, kidney, liver, lung, lymph node, ovary, prostate, skeletal muscle, testes, thyroid, white blood cell) and several embryonic stages (oocyte, pronuclei, zygote, 2-cell, 4-cell, 8-cell, Morula, blastocyst) (Göke et al., 2015). Altogether, the idea is that HERVs are repressed through development but are switched on at specific time point to possibly exert a function. Such function is clearer for some HERVs such as HERV-H in the pluripotency of embryonic stem cells (Lu et al., 2014), and HERV-W in the formation of the placenta (Lavialle et al., 2013). Scientists are still trying to determine whether HERV-K (HML2) is co-opted for a biological role or is on the process of being fully inactivated, since it is the most recently integrated in humans. The ratio between H3K9me3 and H3K4me3, linked to repressed transcription and active transcription state, respectively, was found to be higher for those HERVs most recently integrated (Brattås et al., 2017). They are also more likely to be able to replicate, a viral feature that needs to be repressed by the host.

The active profile of some HERV promoters, such as HERV-H and HERV-L that are older than HERV-K (HML2) and are activated at specific stages of embryogenesis, suggests that during evolution, endogenized retroviruses go through a step of repression first, followed by a step of co-option in case they can provide benefits. Even when co-opted, their window of action seems narrowed in a lifetime for some sort of balance.

5.6.2 HERV-K (HML2)-associated transcription factor

Downstream of methylation, transcription factors are required to modulate transcription. Lines of evidence suggest the requirement of transcription factors, as unmethylated LTR promoters cloned in expressing vectors, introduced in different cell lines can show activity or high transcription levels in one, and no activity or low levels of transcription in the second (Fuchs et al., 2011; St Laurent et al., 2013). Also, while reprogramming or differentiation of cells *in vitro*, HERV-K (HML2) transcription is 'reactivated' or lost, respectively, alongside with expression or loss of transcription factor such as OCT4, SOX2, NANOG (Fuchs et al., 2013). Furthermore, knock-out of OCT4 and SOX2 for HERV-K (HML2); and OCT4, NANOG, LBP9 for HERV-H results in the decrease transcription of the corresponding HERV lineage (Grow et al., 2015; Wang et al., 2014). OCT4, SOX2, NANOG, LBP9 binding sites were found to belong to or to be close to LTR regions (Grow et al., 2015; Santoni et al., 2012; Wang et al., 2014).

Regarding HERV-K (HML2), an exhaustive map of transcription factors potentially able to bind to LTRs was proposed by Manghera et al. (Manghera and Douville, 2013). Among all the transcription factors suggested, the only ones that have been tested experimentally to bind to LTR using chromatin immunoprecipitation (ChIP) and electrophoretic mobility shift assay (EMSA), are the following: OCT4 (Grow et al., 2015), NF-κB and NF-AT (Gonzalez-Hernandez et al., 2012), MITF-M (Katoh et al., 2011), Sp1 and Sp3 (Fuchs et al., 2011), YY1 (Knössl et al., 1999).

5.7 Role in physiology

5.7.1 LTRs fine-tune gene expression

It is believed that LTRs accumulating across the genome provide a more complex gene control and may be involved in species evolution and differentiation, as LTRs provide promoters in genome regions that are non-conserved among species (Bourque et al., 2008). It should also be noted that ERVs vary between animals, lineages can be specific to one species or shared between several species, e.g. HERV-K (HML2) is not human-specific and is found in many primates (reviewed in Hanke et al. (2016)). LTRs could help dissemination of transcription factor binding sites across the genome, they are associated with transcriptional activity of genes in the vicinity (within 50-100kb); and knock-out of KAP1 results in an increase of their transcription (Brattås et al., 2017; St Laurent et al., 2013). Although, most transcripts within the vicinity of HERVs are long non-coding and very long non-coding RNA (lncRNA and vlncRNA), likely associated with epigenetic regulation (Lu et al., 2014; St Laurent et al., 2013). The high frequency of such RNAs in close vicinity of LTRs are likely due to their location, usually in region without genes known as "desert regions" (St Laurent et al., 2013). Furthermore, HERV-H lncRNA was shown to be almost exclusively nuclear and to bind to proteins such as OCT4, p300, MED, CDK8. Also, the knock-down of HERV-H transcripts using shRNAs efficiently results in the loss of pluripotency. Altogether, it suggests that ERVs are involved in pluripotency and fine-tuning of gene expression (Figure 7) (Lu et al., 2014).

Regarding HERV-K (HML2), it is not clear whether LTRs are responsible for such cell pluripotency, even if levels of HERV-K (HML2) transcripts correlates with the expression of OCT4 and NANOG, along with the loss of LMNA expression, a marker of differentiated fibroblasts (Fuchs et al., 2013). The most recent study by Grow et al. (2015) suggests a role of HERV-K (HML2) in pluripotency, as they found high levels of transcription in human embryonic stem cells (hESC) (≥ 10 RKPM); along with binding to p300 complex, as for HERV-H (see Figure 7). HERV-K (HML2) transcription was previously reported at lower levels in hESC (0.5-2 RKPM), using RNAseq (Göke et al., 2015). Nonetheless, both studies found a significant increase of HERV-K (HML2) in hESC from the stage 8-cell, and morula of embryos, in comparison to the other stages (Göke et al., 2015; Grow et al., 2015). Further down the

line, the rise of HERV-K (HML2) expression was linked to the induction of anti-viral mechanisms that could limit potential *in utero* infection by exogenous viruses (Grow et al., 2015).

5.7.2 HERV-derived syncytins play a role in placentation

At least seven different syncytins have been discovered in mammals, each from independent endogenization of different ERVs in different species (Lavialle et al., 2013). In human, *syncytin-1* and *syncytin-2* are derived from a member of the ERV lineages HERV-W and HERV-FRD, respectively (Blond et al., 2000; Mi et al., 2000). Each gene has spent a relatively long time in the host genome (more than 30 million years) since the retroviral integration, but exhibits an intact ORF with a low levels of polymorphism in the human population, in comparison to more recently integrated loci such as some HERV-K (HML2) *env* (less than 5 million years) (de Parseval et al., 2005), suggestion that such retroviral *env* genes have been co-opted. The syncytins and their receptors are expressed in the placenta (Blond et al., 2000; Esnault et al., 2008; Malassiné et al., 2005; Muir et al., 2006). In addition, the involvement of syncytins in the development of the placenta was demonstrated *in vitro* as *syncytin-1* and *syncytin-2* mediate cell-cell fusion (Blaise et al., 2003; Blond et al., 2000; Mi et al., 2000) and the knock-down of syncytins in trophoblast-derived cell lines abolishes such fusion (Frendo et al., 2003; Vargas et al., 2009). The ultimate line of evidence was produced in the mouse. The mouse genome possess two genes encoding retroviral Env proteins termed *syncytin A* and *B* that are distinct from the human syncytins, but share the same characteristics: their expression is placenta-specific and they promote cell-cell fusion (Dupressoir et al. 2005). in fact, all knock-out mice for the genes encoding syncytin-A die during gestation, while knock-out of *syncytin-B* results in growth retardation and a slight decrease in the number of neonates (Dupressoir et al., 2011, 2009).

Regarding HERV-K (HML2), even though the Env is expressed in the placenta (Kämmerer et al., 2011) and possesses a fusogenic ability (Dewannieux et al., 2005), there is no evidence of HERV-K (HML2) Env being involved in placenta development. Moreover, HERV-K (HML2) *env* is more recently integrated but is more polymorphic in the human population, along with a higher rate of non-synonymous mutations, in comparison to the less recently integrated *syncytin-1* and *syncytin-2* (de Parseval et al.,

2005). This suggests that, at this stage in evolution, HERV-K (HML2) *env* is likely not co-opted as the syncytins are.

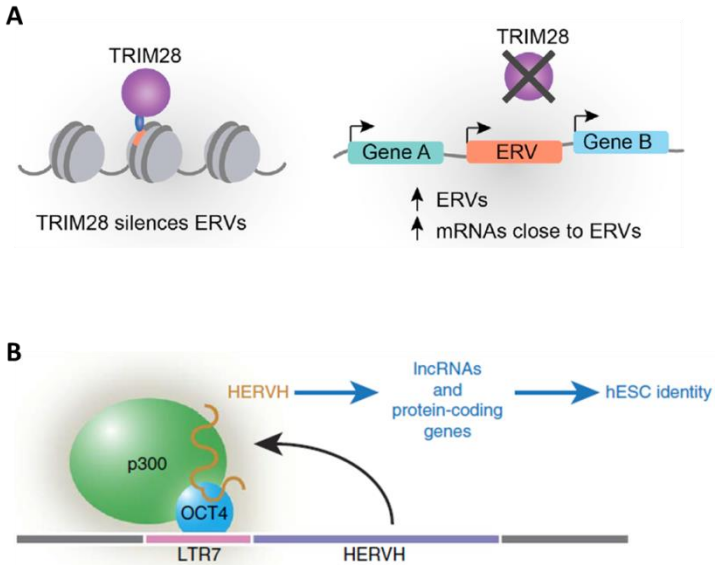


Figure 7 : HERV-driven gene fine-tuning.

(A) TRIM28/KAP1 binds to ERV LTRs across the genome. TRIM28/KAP1 recruits enzymes with DNA methylation activity, resulting in a silencing of transcription from LTR and from genes in the vicinity. Knocking out TRIM28/KAP1 results in an activation of transcription from LTR and from genes in the vicinity. Adapted from (Brattås et al., 2017). (B) HERV-H loci are usually located in gene-free region. Transcription from HERV-H LTRs (LTR7) produces long-non-coding RNAs that remain in the nucleus and associate with chromatin-modulating and transcription complexes such as p300. Knocking down HERV-H RNAs in human embryonic stem cell (hESC) results in the loss of pluripotency. Adapted from (Lu et al., 2014).

5.8 HERV-K (HML2) in disease conditions

5.8.1 HERV-K (HML2) in cancer

The main link between HERV-K (HML2) and cancer is the fact that HERV-K (HML2) is generally up-regulated in the context of cancer. The first question to raise is whether HERV-K (HML2) up-regulation is exclusive to cancer. The second is whether HERV-K (HML2) is a cause or a consequence of cancer. The third is whether HERV-K (HML2) proteins are targetable as tumour-associated antigens.

5.8.1.1 HERV-K (HML2) expression in cancer

To date, many reports described the expression of HERV-K (HML2) in cancers and tumours (Buscher, 2005; Contreras-Galindo et al., 2008; Goering et al., 2011; Ishida et al., 2008; Sauter et al., 1995; Wang-Johanning et al., 2008, 2007, 2001), with very few showing no expression of HERV-K (HML2) in the tumours (Kessler et al., 2014). Since HERV-K (HML2) sequences are mainly defective (Subramanian et al., 2011), the presence of HERV-K (HML) proteins appears to be a relevant readout. There is a clear protein overexpression for four types of cancer and tumours, at a significant frequency in patients: breast cancer (expression of HERV-K (HML2) Env was found in 102 of 119 (85.7%) breast cancer tissues, whereas Env expression was found in 4 of 56 (7.14%) normal breast tissues) (Wang-Johanning et al., 2001), melanomas (of the melanoma biopsies investigated, 7 of 15 (47%) were positive for HERV-K (HML2) Env expression and 16 of 23 (70%) were positive for Gag expression) (Buscher, 2005), ovarian cancers (all types of ovarian cancers pulled together, 467 out of 553 (84%) of ovarian tissues were positive for HERV-K (HML2) Env staining, whereas all normal ovarian tissues were negative (0/3)) (Wang-Johanning et al., 2007), and seminomas that affect testes (5 out of 8 (62.5%), as assayed by western blotting and 6 out of 11 (55%) biopsies, as assayed by immunohistochemistry. Tissues from healthy testes were stated by authors as negatives) (Sauter et al., 1995). The plasma of patients who developed tumours was also assessed for the expression of HERV-K (HML2) components. Contreras-Galindo et al. found HERV-K (HML2) viral particles in the plasma of lymphoma patients (Contreras-Galindo et al., 2008).

Regarding transcription, HERV-K (HML2) is not exclusive to cancer. Rec and Np9 transcripts were found in the heart, the brain, the placenta, the lung, the liver, the skim muscle, the kidney, the pancreas, the spleen, the thymus, the prostate, the testis, the ovary, the small intestine, the colon and leukocytes (Schmitt et al., 2015). However, studies comparing expression of HERV-K (HML2) transcripts in tumours and healthy tissues showed an increased expression in cancerous biopsies. It is the case for breast cancer (Wang-Johanning et al., 2001), prostate cancer (Goering et al., 2011; Ishida et al., 2008) and some ovarian cancers (Wang-Johanning et al., 2007).

5.8.1.2 HERV-K (HML2) role in cancer

The main interest in HERV-K (HML-2) lies in cancer research. In fact, HERV-K (HML2) reactivation seems to be a feature of cancer. This is understandable as hypomethylation events and loss of tumour suppressors that usually control genes linked to pluripotency, commonly happen in tumour development (Ehrlich, 2009, 2002). There is some evidence for HERV-K (HML2) playing a role in tumour growth. First, knocking-down of the HERV-K (HML2) Env in breast cancer cell lines (MDA-MB; SKBR3; MCF7; Hs578T) *in vitro* and inoculating them subcutaneously into the flank of immunodeficient nude mice resulted in decrease of the tumour size in the xenografts, with reduced metastasis to the lung (F. Zhou et al., 2016). Also, when HERV-K (HML2) *env* was overexpressed in those breast cancer cell lines and HEK293T, it activated ERK-MAPK pathway, likely contributing to oncogenesis, with increased metastasis to lung but no difference in the tumour size in the xenografts (Lemaître et al., 2017; F. Zhou et al., 2016). Second, accessory proteins Rec and Np9 from HERV-K (HML-2), overexpressed in cell lines, showed an increase in the activation of signaling pathways such as Numb/Notch1; ERK1/2; c-myc/Akt; Wnt/β-catenin networks, along with increase of tumour growth when xenografted in mice (Chen et al., 2013).

5.8.1.3 Anti-HERV-K (HML2)-specific responses in cancerous patients

Some reports described HERV-K (HML2)-specific T-cell responses as well as antibody responses in cancer. Patients with breast cancer (Wang-Johanning et al., 2008), ovarian cancer (Wang-Johanning et al., 2007), prostate cancer (Ishida et al., 2008), teratocarcinomas and seminomas (Boller et al., 1997; Sauter et al., 1995) were shown

to exhibit higher antibody response to HERV-K (HML2) compared to healthy controls. Interestingly, in a patient with teratocarcinoma at the time of diagnosis, the HERV-K (HML2) Gag-specific antibody response was surprisingly stronger than the response recorded after tumour removal. Interestingly, there are more sera from patients that exhibited a positive response to HERV-K (HML2), compared to sera from healthy controls (62.5% vs 11.1%) (Boller et al., 1997). These findings suggested that tumours producing HERV-K (HML2) proteins as antigens, can stimulate antibody responses against HERV-K (HML2). More recently, Johanning et al. made a proof of concept of breast cancer immunotherapy using a HERV-K (HML2)-specific monoclonal antibody. In a mouse model, they showed tumour regression following treatment with the antibody (Wang-Johanning et al., 2012). However, antibody-mediated tumour regression remains to be understood. Similarly, Kraus et al. showed that vaccination against HERV-K (HML2) proteins inhibits tumour growth in a mouse model (Kraus et al., 2014, 2013). Also, it is shown that the use of chimeric antigen receptor T cells (CAR-T cells) specific to HERV-K (HML2) antigens, can result in cancer cell killing *in vitro* and *in vivo* in a mouse model, which consisted of immunodeficient mice inoculated in the flank with breast cancer cell lines (MDA-MB) and that received intravenous infusion of HERV-K (HML2)-specific CAR-T cells days post-tumour injection (Zhou et al., 2015).

5.8.2 HERV-K (HML2) and autoimmunity

Autoimmune diseases are defined by immune reactions against self-components, leading to the damage and impairment of targeted tissues. Such diseases are usually characterized by the presence of antibody or T cell responses against self-antigens, termed autoantigens. The link between ERVs and autoimmunity has been of interest because some HERVs have been associated with autoimmune conditions, especially for multiple sclerosis (MS) (Christensen, 2016; Morandi et al., 2017) and systemic lupus erythematosus (SLE) (Nelson et al., 2014). How could HERV-K (HML2) contribute to autoimmunity? There are three main possibilities:

1 – By promoting an inflammatory environment, necessary to induce an immune response against autoantigens. This is further developed in the next paragraph about stimulation of innate sensors by ERVs (section 5.8.6).

2 – By being an autoantigen itself. Some authors reported the possibility of HERV-K (HML2) being an autoantigen. Anti-HERV-K (HML2) antibodies have been reported in rheumatoid arthritis, osteoarthritis and systemic lupus erythematosus (Freimanis et al., 2010). But whether these antibodies are responsible for an autoimmune reaction is less plausible, since the level of antibodies are not significantly higher compared to antibodies level in healthy donors (Freimanis et al., 2010). Moreover, the titer of HERV-K (HML2) antibodies appeared to be decreased in patients with psoriasis compare to healthy controls. Also, the expression of HERV-K (HML2) *gag*, *pol* and *env* was lower in patients than in controls (Gupta et al., 2014).

3 – By inducing cross reactivity to self-antigens through molecular mimicry. This is more plausible and discussed in the literature. Usually autoantigens are generated from proteins that should be sequestered in specific cellular compartments in homeostasis. Since HERV-K (HML2) can be expressed and potentially secreted outside of the cells, it is likely more targetable by the immune system and could possibly trigger an immune response by molecular mimicry. This phenomenon was described for some pathogens that harbor similar molecular patterns with autoantigens, which induce immune auto-reactivity usually correlated with a history of infection (Oldstone, 2014). In fact, HTLV-related ERV (HRES-1) possesses a peptide (RPPRP) with an identical amino acid sequence to that of a small nuclear ribonucleoprotein (SnRNP), described as being an autoantigen for SLE. Also, HRES-1 possesses a peptide with a predicted structure that is homological to the ribonucleoprotein SmD (EAVAGRGR (HRES-1) – EAGAGRVR (SmD)), another autoantigen in SLE (Nelson et al., 2014). Molecular mimicry between HERV-R (ERV-3) Env and some SLE autoantigens have been demonstrated (Nelson et al., 2014). Do note that antibodies typically recognize a linear epitope of 5-8 amino acids (Elbakri et al., 2010).

5.8.3 Stimulation of innate sensors by endogenous retroelements

Innate sensors designate the cellular receptors that bind pathogen and danger associated molecular patterns (PAMPs and DAMPs), and trigger an inflammatory response. PAMPs and DAMPs can be of different nature. Nucleic acids (RNA, DNA) and

proteins can trigger an inflammatory response through innate receptor-associated cell signaling (Medzhitov, 2001).

Many lines of evidence suggest a link between retroelements and innate sensing. Mutations in the gene *TREX1* result in Aicardi-Goutières syndrome (AGS), a severe autoimmune disease (Crow et al., 2006). TREX1 is an exonuclease that metabolizes single-stranded DNA (ssDNA) (Lindahl et al., 1969). TREX1-deficient mice possess an accumulation of intracellular ssDNAs (Yang et al., 2007). The authors also showed that ssDNAs were associated with the endoplasmic reticulum (ER), since they co-localized with calreticulin, an ER protein. Similar features were observed in primary fibroblasts derived from AGS patients (Yang et al., 2007). The study of ssDNAs revealed more DNA fragments from endogenous retroelements accumulating in TREX1-deficient mice compared to wild type mice (Stetson et al., 2008). Also, the authors demonstrated that TREX1-deficient mice exhibit a higher level of type I interferon (IFN) expression, autoantibodies, and was associated with a higher mortality (Stetson et al., 2008). Such ssDNAs seem to target IFN pathway and trigger autoimmunity with autoantibodies as a feature. However, it is unclear how ssDNAs from endogenous retroelements are accumulated in the cytoplasm. One explanation could be that RNA from ERVs is transcribed in the nucleus and is reverse transcribed into DNA in the cytoplasm, but retroviral RNA is usually converted into double-stranded DNA (dsDNA) during reverse transcription. To investigate the role of retroelement-related RT in the accumulation of ssDNAs in cytoplasm, RT inhibitors could be used in TREX1-deficient mice, followed by ssDNA staining. Also, the transcription level of retroelements RNAs should be monitored since they could be substrates to create dsDNAs and ultimately ssDNAs. Such a concept appears to be plausible since TREX1-deficient mice treated with antiretroviral drugs (ART) nevirapine, emtricitabine and tenofovir exhibit a higher lifespan with a lower inflammation (Beck-Engeser et al., 2011). All three drugs are reverse transcriptase inhibitors used to treat HIV-infected patients.

Also, ERV-derived transcripts were shown to be increased in B cells from mice immunized with a T-independent antigen, resulting in the activation of cGAS-cAMP-GMP-STING pathway, through the binding of ERV transcripts to Retinoic Acid-Induced Gene-1 (RIG-1), another innate sensor (Zeng et al., 2014). The stimulation of

cGAS-cAMP-GMP-STING pathway results in a pro-inflammatory response through the production of IFNβ, TNFα, IL-6 and IL-12, leading to ERV-specific antibody secretion by B cells (Zeng et al., 2014). Also, an impairment of such antibody response was also observed in cGas- and Sting-deficient mice. Interestingly, TREX1-deficient mice exhibit autoantibodies responsible for an autoimmune-like phenotype (Zeng et al., 2014).

Regarding ERV-derived proteins, there are studies suggesting that HERV-W can trigger inflammatory response by stimulating TLRs via freely released envelope glycoproteins (Rolland et al., 2006; Perron et al., 2013, 2001; Saresella et al., 2009; reviewed in Küry et al., 2018). However, the stimulation of peripheral blood mononuclear cells (PBMCs) by a recombinant HERV-K (HML2) Env *in vitro* resulted in the secretion of IL-10, an anti-inflammatory cytokine (Morozov et al., 2013).

More extensively, the fact that ERVs could be PAMPs for the immune system or up-regulated during an immune response stresses their potential involvement during a pathogenic invasion. In this regard, as ERVs share features of other infectious retroviruses, the interest is put on whether there is a crosstalk between ERVs and incoming retroviruses. The next section presents the interactions that could exist between HERV-K (HML2) and HIV as they have been given the most interest during the last decades. In this thesis, we address related questions in a macaque model infected with SIV.

5.8.4 HIV and HERV-K (HML2)

HERV-K (HML2) was found to be upregulated in HIV-infected cells *in vitro* (Contreras-Galindo et al., 2007; Jones et al., 2012), and in PBMCs from seropositive patients (Bhardwaj et al., 2014). In addition, HERV-K (HML2)-specific antibody and T cell responses were observed in the blood from patients as compared to controls (Garrison et al., 2007; Michaud et al., 2014a). The main question is to understand how HIV could interact with HERV-K (HML2). Several aspects of the viral cycle should be considered: The first aspect is whether HERV-K (HML2) could interfere with or facilitate HIV entry. The second aspect concerns the potential involvement of HERV-K (HML2) in the steps that lead to HIV integration, namely: reverse transcription and

integration. The third aspect concerns the potential of both viruses to modulate each other's expression, transcriptionally and translationally. The fourth aspect is the possibility for HERV-K (HML2) PR being involved in the cleavage of HIV polyproteins into mature proteins that assemble into an infectious particle; this step is critical since a protease-defective virus loses infectivity (Kohl et al., 1988). The last aspect concerns the possibility that the active virus (HIV) package the defective one (HERV-K (HML2)) and enable its transmission – i.e. whether HERV-K (HML2) RNA genome can be retrieved in HIV viral particles.

5.8.4.1 HERV-K (HML2) proteins modulation of HIV infection

It is not known whether the presence of HERV-K (HML2) in human cells modulates HIV infection (i.e. increase or decrease the susceptibility to HIV infection). There is no evidence of individuals harboring specific HERV-K (HML2) loci that correlate with higher or lower levels of HIV viremia.

5.8.4.2 HERV-K (HML2) complementing HIV reverse transcription and integration

There is no evidence concerning HERV-K (HML2) complementing HIV reverse transcription. We do not know whether HERV-K (HML2) RT can convert HIV RNA into DNA during HIV infection. On another hand, one group was interested in the possible complementation of HIV integration by HERV-K (HML2) integrase (Ogata et al., 1999). The use of IN-defective HIV supplemented with HERV-K (HML2) IN resulted in a massive loss of infectivity. This finding suggests HERV-K (HML2) cannot efficiently complement HIV integration.

5.8.4.3 HIV modulates HERV-K (HML2) expression

The third part concerns the ability of HIV to modulate HERV-K (HML2) virus expression and vice versa. Whether, HERV-K (HML2) can upregulate HIV expression has never been investigated. It is probably because such experiment would require infectious HERV-K (HML2) particles and latently HIV-infected cells. However, HERV-K (HML2) is mostly defective and infectious particles are affordable only by the reconstitution of a viable HERV-K (HML2) replication-competent provirus

sequence (Dewannieux et al., 2006; Young and Bieniasz, 2007), which could happen theoretically by a recombination of loci harboring intact and functional ORFs, but it is not the case *in vivo* as such HERV-K (HML2) recombinant has never been reported (Subramanian et al., 2011). On the contrary, several groups were able to show HERV-K (HML2) upregulation both *in vivo* and *in vitro* during HIV infection (Bhardwaj et al., 2014; Contreras-Galindo et al., 2012, 2007; Jones et al., 2012). Even if there is a debate about the magnitude on such phenomenon, more transcripts from HERV-K (HML2) were found in PBMC of HIV seropositive patients compared to uninfected individuals (Bhardwaj et al., 2014). Also, seropositive patients exhibited a higher specific immune response to HERV-K (HML2) compared to uninfected controls (Michaud et al., 2014a). Such HERV-K (HML2)-specific immune response underlies the possible production of HERV-K (HML2) proteins to provide antigens. Furthermore, HIV Tat activates the NF-κB pathway that can lead to the binding of the transcription factor on HERV-K (HML2) LTR and promote transcription (Gonzalez-Hernandez et al., 2012), this result suggests a mechanism for HIV-driven HERV-K (HML2) transactivation.

5.8.4.4 Cleavage of viral polyproteins

HERV-K (HML2) protease (PR) was shown to cleave HIV Gag polyprotein. However, it appeared to happen on non-canonical cleavage sites, leading to different sub-products; and PR-defective HIV supplemented with HERV-K (HML2) PR loses its infectivity (Padow et al., 2000).

5.8.4.5 Viral packaging and pseudotyped HIV-HERV-K (HML2) viruses

The last part concerns the possibility of co-packaging of HIV and HERV-K (HML2), and whether HERV-K (HML2) facilitates the release of HIV viral particles. Firstly, the co-packaging of viral genome is a fundamental question to address, since it could be a proof of concept of HERV-K (HML2) dissemination in the human genome. If HERV-K (HML2) is packaged in HIV virions, it could be passed more efficiently from cell to cell, since HERV-K (HML2) is not infectious. Zeilfelder et al. addressed that question, and succeeded artificial packaging of HERV-K (HML2) into HIV particles (Zeilfelder et al., 2007).Whether this HERV-K (HML2) sequences are further integrated in the target cells and the relevance *in vivo* need to be assessed. It would be interesting to

specifically immunoprecipitate HIV virions from the plasma of seropositive patients using an anti-HIV antibody, and perform a reverse transcriptase PCR to identify HERV-K (HML2) genomic RNA, then confirm whether HERV-K (HML2) is packaged in HIV particles *in vivo*.

Secondly, regarding the possibility of HERV-K (HML2) facilitating HIV virion release, the viral release is limited by the activity of Tetherin, which retains newly formed virions on the cell surface (Kuhl et al., 2011). It was recently shown that HERV-K (HML2) can antagonize the antiviral activity of Tetherin (Lemaitre et al., 2014). Such finding underlines the possibility of HERV-K (HML2) to facilitate HIV virion release.

5.8.4.6 Anti-HERV-K (HML2)-specific responses in HIV-positive patients

Our immune system involves both innate and adaptive responses against pathogens. Innate responses are triggered early during a pathogen invasion. The role of innate immunity is to recognize the danger through Pathogen Recognition Receptors (PRR, innate sensing), and fight against the early burden of the pathogen through soluble factors (complement, cytokines, chemokines, inflammation), innate immune cells (Macrophages, Dendritic cells, Polynuclear Neutrophils, Eosinophils, NK cells), and the ultimate triggering of the adaptive immune response (antibodies, B cells, T cells).

The adaptive immune response is mainly composed by a humoral response through antigen-specific antibodies, and a cellular response mediated by antigen-specific T lymphocytes. There are evidence of HERV-K (HML2)-specific T cell responses in PBMCs from HIV-positive individuals (Garrison et al., 2007). In addition, HERV-K (HML2)-specific CD8+ T cell clones can kill HIV-infected cells *in vitro* (Jones et al., 2012). A significant higher antibody titer against HERV-K (HML2) TM were shown to be present in seropositive patients compared to uninfected individuals (Michaud et al., 2014a). Interestingly, there was not a significant difference between the titer of anti-HERV-K (HML2) SU antibodies titer between the seropositive and the seronegative groups. Such anti-HERV-K (HML2) TM-specific antibody can induce HIV-infected cell killing through Antibody Dependent Cell-mediated Cytotoxicity (ADCC) (Michaud et al., 2014b). All together, these findings suggested that HERV-K (HML2) antigens (protein or polypeptides) are generated during HIV infection. In this case,

HERV-K (HML2) provides targeted antigens on HIV-infected cells, indirectly sustaining an anti-HIV immune response.

Similarly, in the macaque, we are asking whether SIV can modulate ERV expression, and whether such expression of ERVs could be involved in the immune response to SIV.

5.9 Merlin-deficient tumours: the schwannomas.

Merlin is a protein from the family of Ezrin Radixin Moesin superfamily and is encoded by the *NF2 gene* located on chromosome 22 (reviewed in Bretscher et al., 2002). The alteration of the *NF2 gene* causes neurofibromatosis type 2 (NF2) (Rouleau et al., 1993; Trofatter et al., 1993), a condition defined by the presence of schwannomas; they are benign tumours arising from Schwann cells that myelinated nerve fibers in the nerves (Bhatheja and Field, 2006). In fact the presence of bilateral vestibular schwannomas (found in the vestibular nerve linking the inner ear to the brain) is a criteria for the diagnostic of NF2. Such genetic condition can be inherited in an autosomal dominant fashion, or arise spontaneously (sporadic cases). Recently, it has been observed that an alteration of the *NF2 gene* encoding Merlin account for 70% of schwannomas that occur sporadically (Agnihotri et al., 2016).

At the cellular and molecular level, schwannomas are characterised by an increased cell division linked to a high expression of cyclin D1, many growth factor receptors and the constitutive activation of pathways such as the Ras-Mek-Erk and Pi3K-Akt pathways marked by the constitutive phosphorylation of the signal proteins ERK (p-ERK), AKT (p-AKT) and FAK (p-FAK) (Figure 8) (reviewed in Ammoun and Hanemann, 2011; Hilton and Hanemann, 2014).

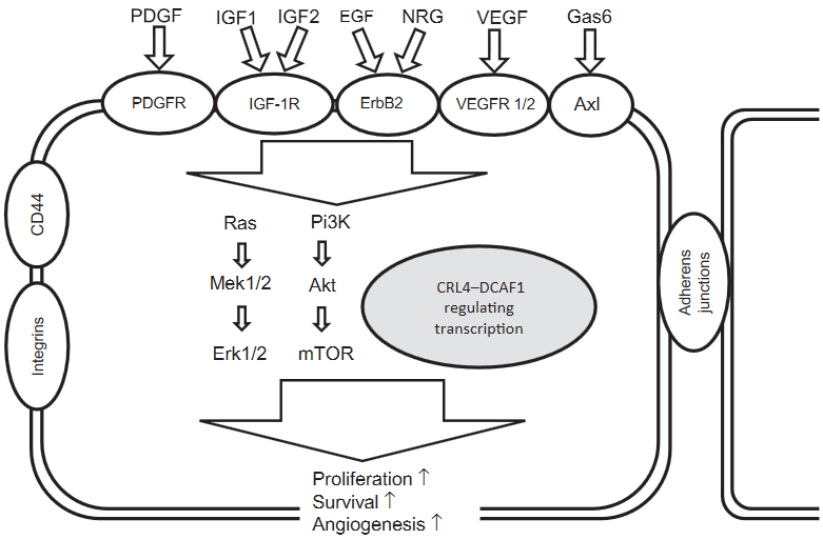


Figure 8: Merlin loss-induced pathways in schwannomas.

In the absence of Merlin, the CRL4–DCAF1 complex is active, resulting in increased expression of a number of genes, including integrins and growth factor receptors. Merlin-deficient schwannomas are hence more sensible to growth factor-induced proliferation and survival linked to an activation of the ERK and AKT pathways. EGF = epidermal growth factor; ErbB2 = epidermal growth factor receptor 2; Gas 6 = growth arrest specific 6; IGFR = insulin-like growth factor receptor; NRG = neuregulin; PDGFR = platelet-derived growth factor receptor; VEGFR = vascular endothelial growth factor receptor. Figure from Hilton and Hanneman (2014).

5.10 The Hippo pathway

The Hippo pathway was first described in *Drosophila*. Later, homolog proteins of the core cascade of the pathway were validated in mammals (mainly using mice) (Pan, 2010). Such signal cascade is composed by the following effectors: MST1 and 2, SAV1, LATS1 and 2, MOB1A and 1B, YAP and TAZ, the transcription factors TEAD1-4. Physiologically, MST1/2 phosphorylates LATS1/2, which phosphorylates YAP/TAZ, which in a phosphorylated form cannot bind to TEAD. If not in complex with YAP/TAZ, TEAD does not bind to DNA, avoiding the transactivation of genes that are anti-apoptotic and pro-proliferation (Figure 9) (Juan and Hong, 2016). The Hippo pathway is then relevant to control organ size, in fact over-expression of YAP in mouse liver induced an increase in the organ size with similar features to human colon cancers (Camargo et al., 2007).

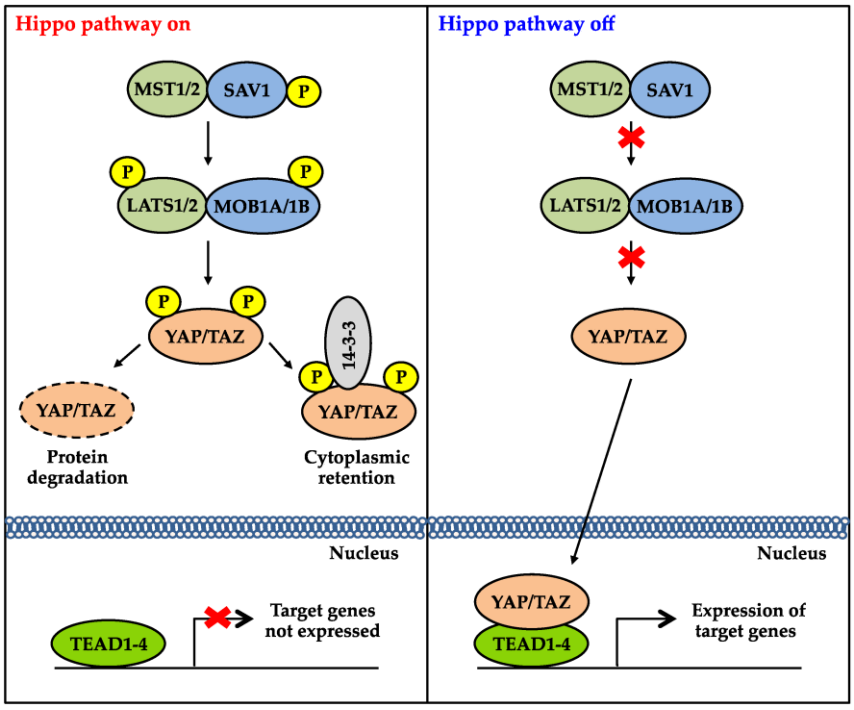


Figure 9: The core signaling cascade of the Hippo pathway.

When the hippo pathway is on: MST1/2-SAV1 complex phosphorylates LATS1/2-MOB1 complex, which phosphorylates YAP/TAZ, such phosphorylation prevent YAP/TAZ translocation and binding to TEAD. When the pathway is off: YAP/TAZ is not repressed by phosphorylation and can translocate inside the nucleus where it binds to TEAD, the YAP/TAZ-TEAD transcription complex induce the expression of many target genes. The figure is from Juan and Hong (2016).

The alteration of the Hippo pathway has been linked to the loss of Merlin and the development of schwannoma, involving a new regulator of the pathway: CTLR4-DCAF1. CTLR4-DCAF1 is an E3 ubiquitin ligase that is inhibited by Merlin. It is suggested that in the absence of Merlin, CRL4-DCAF1 inhibits LATS 1/2 that normally phosphorylates YAP/TAZ. When dephosphorylated, YAP/TAZ binds to TEAD and activate the transcription of oncogenic genes (Figure 10) (reviewed in W. Li et al., 2014).

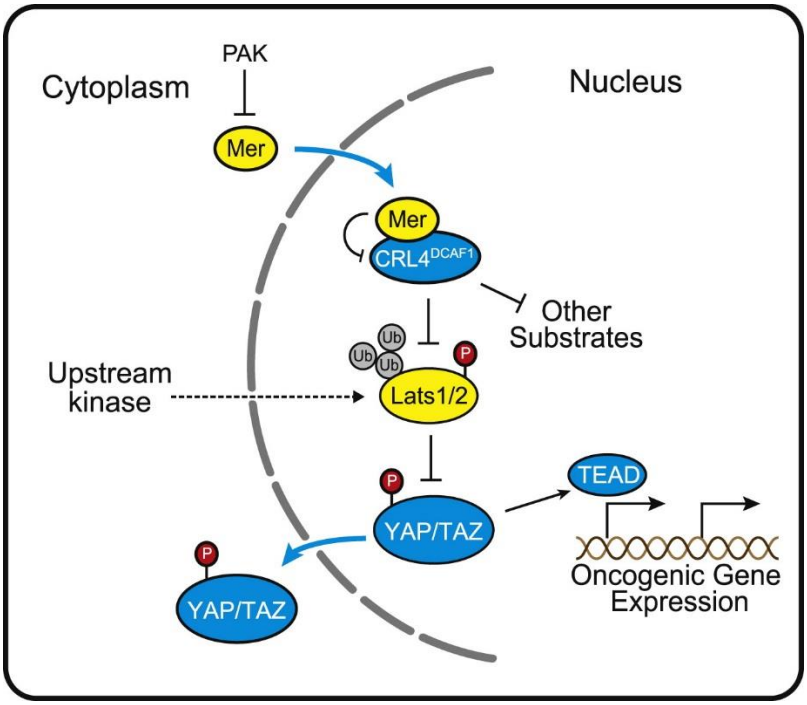


Figure 10: The role of Merlin and CRL4-DCFA1 on the Hippo pathway.

Merlin represses CRL4-DCAF1 activity. CRL4-DCAF1 inhibits LATS1/2 via ubiquitination. LATS1/2 inhibits YAP/TAZ via its phosphorylation. YAP/TAZ binds to TEAD to promote transcription of target genes. Hence, in the absence of Merlin, CRL4-DCAF1 can repress LATS1/2 activity on YAP/TAZ, the latter is free to bind to TEAD and induce expression of oncogenic gene. The figure is from Wei et al. (2014).

5.11 Research challenge

Regarding knowledge on ERVs and diseases, three main points need to be addressed:

First, the need to assess HERV-K (HML2) expression in a wider range of tumours and compared to their healthy counterpart. This could build an exhaustive list of tumours which sustainably express HERV-K (HML2), especially at the protein level. It is well admitted that HERV-K (HML2) is expressed in tumours, however, there is still a need to investigate other tumours, for example, HERV-K (HML2) transcripts could not be detected in astrocytoma and glioblastoma (Kessler et al., 2014). Also, measuring expression on the "healthy" tissues will be very fundamentally useful and could provide a basic knowledge of tissues that would unwittingly be targeted in potential monoclonal antibody-based therapy.

The second point concerns ERV-based therapies. Indeed, as HERV-K (HML2) is expressed in many tumours, the idea of targeting specific tumours by HERV-K (HML2)-specific antibodies could be an appealing strategy. So, it is imperative to test whether ERV-based therapies (monoclonal antibodies, antiretroviral drugs) affect tumours but not healthy tissues.

The third point concerns the role played by ERVs in human and other species. More precisely, LTRs, viral proteins and transcripts need to be assessed for their requirements or not in biological processes, such as cell division, pluripotency, immunity, and tumorigenesis. Lately, scientists gained some insights about it, however, the knowledge mainly concerns the study of HERV-W and HERV-H families in humans.

Finally, on a fundamental basis, transcription factors that drive HERV-K (HML2) expression are to be unraveled. Few of them are shown to bind to the LTRs, but still need to be clearly and further addressed. Indeed, Sp1 binding to GC boxes is believed to promote transcription, however Gonzalez-Hernandez et al. showed it has no effect on transcription using site mutagenesis (Gonzalez-Hernandez et al., 2012).

6 OBJECTIVES

Regarding the research challenges and the materials available in our laboratory, I proposed the following points as objectives to reach in this thesis and beyond.

Objective 1

To test up-regulation of HERV-K (HML2) in our tumour of interest (Section 8.1), along with their healthy counterpart – schwannomas, which arise from Schwann cells.

Objective 2

To test potential therapies, such as anti-HERV-K (HML2) monoclonal antibodies and antiretroviral drugs modelled by the ritonavir – an inhibitor of retroviral proteases (Section 8.1).

Objective 3

To determine whether HERV-K (HML2) contributes to tumour development (Section 8.2). The knock down of HERV-K (HML2) was attempted using RNA interference. Also, because I observed that anti-HERV-K (HML2) monoclonal antibodies could reduce tumour cell proliferation (Objective 2), mechanisms for HERV-K (HML2) release and the possibility of pro-tumoral effect of HERV-K (HML2)-secreted proteins were explored.

Objective 4

To investigate possible mechanics for HERV-K (HML2) up-regulation (Section 8.3). I am particularly interested in TEAD as novel potential transcription factor driving HERV-K (HML2) transcription from LTRs.

Objective 5

To investigate the correlation between a recently integrated ERV in macaques and interferon stimulated genes (ISGs) during viral insult modeled by SIV infection (Section 8.4). My hypothesis is that SIV-infected individuals that harbor higher levels of ERVs should trigger higher interferon response as measured by ISGs.

7 MATERIAL & METHODS

7.1 Ethical approval

Schwann cells and schwannomas were obtained from patients and donors in agreement with an informed consent. The study was granted full national ethics approval by the South West research ethics committee (REC No: 14/SW/0119; IRAS project ID: 153351) and local research and development approval (Plymouth Hospitals NHS Trust: R&D No: 14/P/056 and North Bristol NHS Trust: R&D No: 3458).

Samples from macaques used in the study are archived materials from past vaccine studies for which the NIBSC obtained ethical approval.

7.2 Cell Culture

MCF-7 (Breast adenocarcinoma cell line), HEK 293T were grown in Dulbecco's modified Eagle medium (DMEM, Thermofisher, Cat # 41965062) containing 10% foetal bovine serum (FBS), penicillin/streptomycin (100U/ml) (Thermofisher, Cat # 15140122). HIB (Human Malignant Mesothelioma cell line, Merlin positive), TRA (Human Malignant Mesothelioma cell line, Merlin negative) was cultured in RPMI containing 10% foetal bovine serum (FBS), penicillin/streptomycin (100U/ml) (Utermark et al., 2003). All cell lines were culture in a humidified atmosphere with 5% CO2.

Human primary schwannoma cells (NF2-/-) were isolated as previously described (Rosenbaum et al., 1998). Briefly, after surgical removal of the tumour from patient, Schwannomas were pre-incubated upon arrival for 1-2 days in DMEM with 10% FBS and 100U/ml penicillin/streptomycin (Gibco) in a humidified atmosphere with 10% CO2 and then dissected into small pieces in DMEM with 10% FBS containing 50U/ml penicillin/streptomycin, 160U/ml collagenase type I (Sigma), and 1.25U/ml dispase grade I. Tissue pieces were incubated in proteolytic enzymes for 24 h before they were completely dissociated by trituration with a narrowed Pasteur pipette. Cell suspension was harvested and resuspended in culture medium.

Human primary Schwann cells (NF2+/+) were isolated based on a method described previously (Hanemann et al., 1998) with few adjustments. Briefly, peripheral nerves

were obtained from donor. Upon arrival, the nerve was stripped and individual fascicles were removed. Fascicles were incubated in DMEM with 10% FBS and 100U/ml Penicillin/Streptomycin in a humidified atmosphere with 10% CO2 for 1-2 days. Then fascicles were trimmed into 1-mm-long pieces in DMEM with 10%FBS containing 160U/ml collagenase type IIA and 0.8 U/ml dispase grade I. Fascicles were incubated in proteolytic enzymes overnight before they were completely dissociated by trituration with a narrowed Pasteur pipette. Cells were harvested and resuspended in culture medium.

Human primary schwannoma (NF2-/-) and Schwann cells (NF2+/+) were grown in 6-well plates or 8-chamber Lab-Teks pre-coated with 1 mg/ml poly-L-lysine (Sigma, Cat # P9155) and mouse laminin (Gibco, Cat # 23017015) in a humidified atmosphere with 10% CO2. The culture medium was made of DMEM containing 10% FBS, penicillin/streptomycin (100U/ml), amphotericin B (2.5µg/ml, Thermofisher, Cat # 15290026), insulin (2.5µg/ml, Thermofisher, Cat # A11382II), forskolin (0.5µM, Tocris, Cat # 1099), β-Heregulin (10nM, Bio-Techne, Cat # 396-HB), without (schwannoma) or with (Schwann cells) 3-isobutyl-1-methylxanthine (0.5mM, IBMX, Sigma, Cat # I5879-1G), respectively.

7.3 Antibodies

Mouse anti-HERV-K (HML2) Env (AMSBio, Cat # HERM-1811-5), anti-HERV-K (HML2) capsid (AMSBio, Cat # HERM-1831-5), anti-HERV-K (HML2) Gag (AMSBio, Cat # HERM-1841-5) was used at a dilution of 1:250; 1:500 or 1:1000. Rabbit anti-phospho ERK (Promega, Cat # V803A) was used at a dilution of 1:5000. Rabbit anti-phospho AKT (New England Bioloabs, Cat # 9271) at a dilution of 1:500. Rabbit anti-phospho FAK^{Y397} (New England Bioloabs, Cat # 3283) was used at a dilution of 1:500. Rabbit anti-ERK (New England Biolabs, Cat # 4695), anti-AKT (New England Biolabs, Cat # 4691), anti-FAK (New England Biolabs, Cat # 3285), anti-Cyclin D1 (New England Biolabs, Cat # 2922) were used at a dilution of 1:500. Rabbit anti-DCAF1 (also known as VPRBP, Proteintech, Cat #11612-1-AP) and anti-CTGF (Abcam, Cat # ab6992) were used at a dilution of 1:1000. Secondary Goat anti-mouse and anti-rabbit coupled to HRP were used at dilution 1:20 000 or 1:10 000. Mouse anti-CD63 (Thermofisher, Cat # 10628D) was used at a dilution of 1:500.

Secondary HRP-conjugated goat anti-mouse (Biorad, Cat # 172-1011) and goat anti-rabbit (Biorad, Cat # 172-1019) were both used at a dilution of 1:10 000.
Rabbit anti-c-Jun (New England Biolabs, Cat # 9165), mouse anti-Ki67 (Agilent Technologies, Cat # M7240) was used at a concentration of 1:100. Secondary goat anti-mouse coupled to Alexafluor 488 (Thermofisher, Cat # A11001) and 594 (Thermofisher, Cat # A11005); and goat anti-rabbit coupled with Alexafluor 488 (Thermofisher, Cat # A11008) and 568 (Thermofisher, Cat # A11011) were prepared in 1% BSA in PBS at a dilution of 1:500. (Table provided in section 12.5, Table S2).

7.4 Drug treatment

Cells were plated in 6-well plate or 8-chamber Lab-Tek. Ritonavir (Sigma, Cat # SML0491) was used at 0.1 and 1μM for 24-72 hours. The cells were incubated in growth medium with the appropriate concentration of ritonavir for 72 hours. Then the cells were lysed and run in SDS-PAGE. Verteporfin (Tocris, Cat # 5305), inhibitor of YAP-TEAD interaction, was used at a final concentration of 4μg/ml, 4μg were added once, each day for 48 hours until lysis, in a culture volume 2ml of medium. Anti-HERV-K (HML2) capsid and Env were used at 1μg/ml in culture, normal mouse IgG control (Santa Cruz, Cat # SC-2025) were used at 1μg/ml dilution in culture.

7.5 HERV-K 9-mer peptides

The peptides used in schwannoma culture were HERV-K (HML2) Env-derived QIFEASKAHL peptide (IBA Lifesciences, Cat # 6-7072-901) and HERV-K (HML2) Gag-derived VMAQSTQNV peptide (IBA Lifesciences, Cat # 6-7073-901). Both were used at a concentration of 5μg/ml for 72 hours in non-supplemented DMEM. Cells were starved in non-supplemented DMEM for 24 hours, prior to the incubation with HERV-K (HML2) peptides.

7.6 Plasmids

pBluescript II SK+ (pBS, 3.0kb) was purchased from Stratagene (Cat # 212205; 1μg/μL). pBS-PcEV, was obtained separately by inserting PcEV *pol* amplicon into the multiple cloning site (MCS) of a pBS vector. PcEV *pol* amplicon was inserted into ApaI (nt 604) restriction sites. This was achieved by GENEWIZ, Inc. Briefly, the amplicon

was synthtised (as part of their gene synthesis service), cloned into the vector purchased pBS vector (Stratagene), and the resulting vector was sequenced as a quality control (the sequence of the entire vector + amplicon is provided in supplementary information, section 12.5).

pGFP-V-RS (7584bp) and pGFP-C-shLenti (8.7kb) was purchased from Origene. HERV-K shRNAs were inserted as following: HERV-K (HML2) target sequence-TCAAGAG loop-target sequence reverse complement, by Origene as part of the customer service.

pLenti-H1-shRNA-Rsv was purchased from AMSBio, and HERV-K (HML2) shRNAs were inserted as following: HERV-K (HML2) target sequence-CGAG loop-target sequence reverse complement, by AMSBio as part of the customer service. Each target sequence is provided in a table in paragraph shRNA target sequences (section 7.7).

7.7 shRNA target sequences

Table 2 : shRNA target sequences

target position (nt)	target sequence	ID #	ORF location	Origin (Company)	Reference provirus
Negative control	GTCTCCACGCGCAGTACATTT	scr	-	AMSBio	-
65-87	GACTCCATTTTGTTATGTATTAA	1	HERV-K (HML2) LTR	AMSBio	K113
69-91	CCATTTTGTTATGTATTAAGAAA	2	HERV-K (HML2) LTR	AMSBio	K113
75-97	TGTTATGTATTAAGAAAAATTCT	3	HERV-K (HML2) LTR	AMSBio	K113
350-372	CCATGTGATAGTCTGAAATATGG	4	HERV-K (HML2) LTR	AMSBio	K113
604-626	GGCAGCAATACTGCTTTGTAAAG	5	HERV-K (HML2) LTR	AMSBio	K113
861-883	CCTTATTTCTTTCTCTATACTTT	6	HERV-K (HML2) LTR	AMSBio	K113
878-900	TACTTTGTCTCTGTGTCTTTTTC	7	HERV-K (HML2) LTR	AMSBio	K113
888-910	CTGTGTCTTTTTCTTTTCCAAAT	8	HERV-K (HML2) LTR	AMSBio	K113
890-912	GTGTCTTTTTCTTTTCCAAATCT	9	HERV-K (HML2) LTR	AMSBio	K113
892-914	GTCTTTTTCTTTTCCAAATCTCT	10	HERV-K (HML2) LTR	AMSBio	K113
1114-1136	GGGCAAACTAAAAGTAAAATTAA	11	HERV-K (HML2) Gag	AMSBio	K113
1115-1137	GGCAAACTAAAAGTAAAATTAAA	12	HERV-K (HML2) Gag	AMSBio	K113
1190-1212	GAGTTAAAGTATCTACAAAAAAT	13	HERV-K (HML2) Gag	AMSBio	K113
2306-2328	GTCAAAATTGGAGTACTATTAGT	14	HERV-K (HML2) Gag	AMSBio	K113
2536-2558	GTGGAGTTAATGGCATATGAAAA	15	HERV-K (HML2) Gag	AMSBio	K113
7027-7049	CAGAAGTATATGTTAATGATAGT	16	HERV-K (HML2) Env	AMSBio	K108
7295-7317	TTCTTATCAAAGATCATTAAAAT	17	HERV-K (HML2) Env	AMSBio	K108
7353-7375	CCCAAAGAATCAAAAAATACAGA	18	HERV-K (HML2) Env	AMSBio	K108
7557-7579	GACAAACATAAGCATAAAAAATT	19	HERV-K (HML2) Env	AMSBio	K108
8313-8335	GACTGGAATACGTCAGATTTTTG	20	HERV-K (HML2) Env	AMSBio	K108
Not found	GACACACTGGAGCAGATGTCTTCTACATTGC	e	-	Origene	K108
4905-4933	GCCACTGCACATTCTCCAACAGGCATCAT	f	HERV-K (HML2) Pol	Origene	K108
3329-3354	CCAGTCCAAGAGACAGGATTGCTCAAT	g	HERV-K (HML2) Pro	Origene	K108
6174-6202	AGGAGTGTACCACTCCTCAGATGCAACTT	d	HERV-K (HML2) Pol	Origene	K108
1020-1050	TGGTCATTGAGGACAAGTCGACGAGAGAT	a	HERV-K (HML2) LTR	Origene	K113
949-973	AGGGGCAACCCACCCCTACATCTGG	b	HERV-K (HML2) LTR	Origene	K113
1065-1076//6434-6448**	GTCAGCCTTACGACATTTGAAGTTCTA	c	HERV-K (HML2) LTR/Env**	Origene	K113

**: Spanning splicing sites.

##: Negative control scramble (scr) sequence from Origene is not available.

AY037928.1 and AC072054 are accession numbers for K113 and K108, respectively

For shRNA 1-20, HERV-K113 *gag* and LTR and HERV-K108 *env* were used as query sequences submitted to RNA interference online tool (siDirect version 2.0) which predicted best candidates of siRNA, based on an algorithms designed by (Ui-Tei et al., 2008; Reynolds et al., 2004; Amarzguioui et al., 2004). 5 for each 5 for *gag*, 5 for *env* and 10 for LTRs (5 downstream of the transcription start site, and 5 before upstream of the transcription termination site) were chosen. Each target sequence was synthetised as a shRNA and inserted in lentiviral-expressing vectors (pGFP-C-shLenti) by AMSBio, as a custom service.

shRNAs e,f,g,d was purchased from Origene (Cat # TL313165V) as shRNA-expressing lentiviral vectors using pGFP-C-shLenti vector.

shRNAs a,b,c were chose manually on K113 sequence, and were synthetised and inserted in a pGFP-V-RS shRNA-expressing lentiviral vector by Origene, as a custom service.

7.8 Transfection

On day 1, MCF7 or HEK293T cells were plated in a 6-well plate prior to transfection. On day 2, DNA and MegaTran 1.0 transfection reagent (Origene, Cat # TT200005) was mixed in 200µL of Opti-MEM™ (Thermofisher, Cat # 31985062), and incubate for 15-25 minutes. After that, the mix was added drop by drop in the appropriate well, homogenized by swilling, and the cells were incubated overnight at 37°C. On day 3, the transfecting medium was replaced by normal growth medium. The cells were checked on fluorescent microscope to confirm the transfected state. The cells were then cultured for 3-5 days until lysis and SDS-PAGE.

7.9 Lentiviral production

Origene lentiviruses were provided by the company with an indication of the titer (TU/ml). shDCAF1 lentiviruses was produced and kindly provided by J. L. Rimmer, briefly, on day 1, HEK 293T cells were plated into 100mm dishes. On day 2, cells were transfected with a set of 3 plasmids DNA: shRNA vector, a mix of packaging plasmids that contains pCMV-DR8.2 plasmid encoding for lentiviral core and pVSV-G plasmid encoding for lentiviral Env. On day 3 the medium was replaced with a high-serum medium. On day 4 the medium was collected, replaced with another set of high-serum medium. Harvested medium was centrifuged to remove cellular debris, and the

supernatant containing viral particles was stored at -80°C. On day 5, medium was collected and centrifuged to remove cellular debris, and the supernatant containing viral particles was stored as 2-ml aliquot in cryotubes at -80°C.

7.10 Adenovirus stock production

Merlin (NF2) wild type and control GFP-containing vector adenoviruses (AdV-NF2 and AdV-GFP, respectively) obtained from J. Testa (Xiao et al., 2005) were amplified in 293T cells as previously described (He et al., 1998). Cells were plated in a 500mm dish and culture to a confluence of 80%, then 50µl of previous stock of AdV-NF2 or AdV-GFP were added to the culture. 1-2 days post-infection, medium was removed cautiously. Cells were scraped, harvested in a 15-ml tube and centrifuge for 5 minutes at 2000rpm to discard any remaining medium. Cells were resuspended in 500µl cold PBS and transfer into a 1.5ml Eppendorf tube. A small amount of Liquid Nitrogen was poured into an appropriate container. Using a long pair of tweezers, the small tube was frozen in Liquid Nitrogen and thawed in water bath at 37°C three times, to release AdV virions. The small tube was centrifuge in a bench microfuge to pellet cellular debris. The supernatant containing AdV virions was aliquoted, 50µl per aliquot. Aliquots were stored at -80°C until use.

7.11 Viral Infection

For AdV infection, schwannoma cells were plated and culture to high percentage (~90%) of confluence since AdV produce a lytic infection that induces cell death. 50µl of viral aliquot was introduced in the culture medium. Cells were kept with viruses for 72 hours, or infected medium was changed for fresh medium and kept until day 5 of infection. Then they were lysed for further analysis.

For lentivirus infection, schwannoma and HEK293T cells were plated and cultured to ~70% confluence. Then 2ml of homemade aliquot or the equivalent of 400 000 TU for purchased lentiviruses (Origene, Cat # TL313165V) were introduced in medium, with the addition of 8µg/ml of Polybrene for cell lines or Protamine Sulphate for schwannomas. Cells were kept with viruses for 72 hours, then selected in puromycin (Sigma, Cat # P9620) selection medium (4µg/ml) until desired confluence (>80%).

7.12 Protein lysates

Cells in 6-well plate from culture were washed with PBS two times. Then 50µl of lysing buffer was added in each well. The lysing buffer was RIPA buffer (1mM EDTA: 0.5% sodium deoxycholate (Sigma, Cat # D6750); 0.1% Sodium Dodecyl Sulphate (SDS); 1% Nonidet P40 (Sigma, Cat # 11332473001); 150mM NaCl (Sigma, Cat # S7653); 50mM Tris-HCl pH8)) supplemented with inhibitors of protease (Thermofisher, Cat # 87786), phosphatase B (Santa Cruz, Cat # SC-45045) and C (Santa Cruz, cat # SC-45065). Cell scraper (Fisherbrand, Cat # 11597692) was used to scrape the cells down into the lysing buffer. The cells suspension was collected in 1.5ml Eppendorf tube and incubate on ice for at least 5 minutes. Then the solution was centrifuged at 13 000 rpm for 15 minutes to remove any cellular debris. The supernatant (protein lysate) was collected and stored at -20°C until further use.

7.13 Western Blot Analysis

Cell lysates were mixed with reducing buffer (250mM Tris-HCl pH6.8, 8% SDS, 40% glycerol (Sigma, Cat # G9012) , 200mM dithiothreitol (DTT, Sigma, Cat # 43816) and 0.4% bromophenol blue (Sigma, Cat # B0126)), boiled at 95°C for 5 minutes, and then proteins were separates through a Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) (SDS, Sigma, Cat # L4509). After running, the proteins were transferred from the gel to a PDVF nitrocellulose membrane for 90 minutes (250mA) or overnight (175mA). The membrane was then blocked with Tris Buffer Saline (TBS) containing 1% Tween 20 (Sigma, Cat # P2287) (TBS-T) 5% skim milk powder (Sigma, Cat # 70166), for 60 minutes, to limit non-specific binding. The membrane was then washed 3 times (5 minutes per wash) and incubated with appropriate primary antibody overnight. Visualization of the proteins was achieved using secondary antibodies coupled to horseradish peroxidise (HRP), for 60 minutes incubated at room temperature with enhanced chemiluminescence reagents (ECL™ Western Blotting Detection Reagents, GE Healthcare Life Sciences, Cat # RPN2106) and manual development using autoradiographic Amersham Hyperfilm ECL (GE Healthcare Life Sciences, Cat # 28-9068-36).

7.14 RNA preparation

For studying HERV-K (HML2) transcription, RNA was extracted from cultured cells (MCF7), using an RNEasy kit according to the manufacturer instructions (Qiagen, Cat#74104). RNA was eluted in 50µl and quantitated using a nanodrop (NanoDrop 2000, Thermo Scientific). After DNAse treatment (DNA-free kit, Ambion, Cat # AM1907), each RNA preparation was diluted to 10ng/µl, so that 25ng (2.5µl per reaction) was added in each 25µl qPCR reaction.
For study on macaque PcEV, We analysed acutely SIV-infected macaques from different SIV challenges in historical studies of SIV pathogenesis and vaccination (Berry et al., 2011; Ferguson et al., 2014; Mattiuzzo et al., 2013).

RNA was extracted from 140µl to 1ml of plasmas using a QiaAmp (Qiagen, Cat# 52904) according to the manufacturer's instruction. Each RNA preparation went through DNAse step to remove contaminant DNA. 5µl of DNAse-treated RNA preparation was used per qPCR reaction.

For quantification of cellular-associated transcripts, RNA was extracted from frozen cells, washed once in PBS, and then lysed with guanidine isothiocyanate (Sigma, Cat # 50983). Then 200µl of chloroform was added followed by centrifugation step at 13000 rpm for 2 minutes. The aqueous phase was collected and 1 volume of 100% Ethanol was added, and the samples were loaded onto RNEasy silica column (Qiagen, Cat # 74204). The column was washed and RNA was eluted according to the manufacturer's instruction. RNA was eluted in 50µl and quantitated using nanodrop. Each RNA preparation was diluted to 10ng/µl, so that 50ng (5µl) was added in each qPCR reaction.

7.15 DNAse treatment

Removal of DNA contaminants was conducted using a DNA-free kit (Ambion, Cat # AM1907). 5µl of 10X DNAse Buffer was added to 50µl RNA preparation. Then 1-2µl of TURBO DNAse was added and the solution was incubated for a total of 1 hour at 37°C. Then 10µl Inactivation Reagents were added, and the solution was incubated for 5 minutes flicking the tubes sometimes. The solution was centrifuged at 10 000 x g for

2 minutes to pellet Inactivation Reagents. The supernatant containing RNA was harvested, and used or stored at -80°C for further analysis.

7.16 Primers and probes sequences

Primer sequences for qPCR in human samples were the following, given in 5'-3' orientation:

Table 3 : Primers and probes.

Sequence	Forward	Reverse	Probe
HERV-K(HML2) *pol**	^{4505}AATTGACTGTTATACATTTCTGC	CCGAATCCAATTAATATCTCC^{4695}	
HERV-K(HML2) *env*	^{1017}GCGTGGTCATTGAGGACAAGTC	GGTGCTCGATTGCGGTGTCT^{6496}	
Human GAPDH	^{3059}CTTTTGCAGACCACAGTCCATG	TTTTCTAGACGGCAGGTCAGG^{3274}	
PcEV *pol*	^{3122}CCGTGTCTATCAAGCAATATCC	GGCAGAAGAGGAGTGTTCCAGG^{3227}	^{3197}AACTCGGAGTGTTGCGAC^{3214}
Rhesus GAPDH	^{2268}GGCTGAGAACGGGAAGCTC	AGGGATCTCGCTCCTGGAA^{2314}	^{2288}TCATCAATGGAAGCCCCATCACCA^{2311}
STAT-1	^{443}CAATACCTCGCACAGTGGTTAGAAAA	CGGATGGTGGCAAATGAAAC^{497}	
SIV ltr**	^{641}CTCCACGCTTGCTTGCTTAA	AGGGTCCTAACAGACCAGGG^{761}	^{707}TCCCATATCTCTCCTAGYCGCCGC^{728}
SIV *gag***	^{511}AGTGCCAACAGGCTCAGAAAA	TGCGTGAATGCACCAGATG^{562}	^{533}TTAAAAAGCCTTTATAATACTGTCTGCG^{560}

*: pol primers from Karamitros et al. (2016).

**: SIV primers are based on conserved regions in *gag* (Berry et al., 2008), and in LTR for SIVsmE660 (Berry et al., 2011).

Reference sequences used are: HERV-K (HML2) (K113; NC_022518), human GAPDH (NC_000012.12), PcEV (AF142988.1), rhesus GAPDH (NC_027903.1), STAT-1 (NM_001261614.1), SIVsmE660 ltr (JQ864087.1), SIVmac251 *gag* (KC522253.1). The positions of the nucleotide on the reference sequence annealing with the first nucleotide of the forward primer, the last nucleotide of the reverse primer in the given orientation, the first and the last nucleotides of the probe, are provided.

7.17 qPCR and qRT-PCR

For quantification of cellular-associated HERV-K (HML2) transcripts, each RNA preparation was diluted to 40ng/µl, so that 100ng (2.5µl) was added in each qPCR reaction (20µl). For HERV-K (HML2) *pol*, *env* and human *gapdh*, Sybr green-based qPCR was used. The kit used was RNA-to-Ct Power Sybr green (Applied Biosystem, Cat #4389986). The mix consisted of 6.86µl of nuclease-free water, 1.2µl of each forward and reverse primers (5µM each, final concentration: 300nM), 10µl of 2X

Master Mix (containing SYBR green and polymerase), and 0.16µl of RT-enzyme mix (for reverse transcription).

The qPCR programs were adapted from manufacturer's instructions and were set as follows. The qPCR program for HERV-K (HML2) *pol* (Karamitros et al., 2016) and *gapdh* was as following: 48°C for 30 minutes (reverse transcription step), 95°C for 5 minutes (Holding), 35 cycles of 95°C for 30 seconds, 47°C for 30 seconds and 72°C for 1 minute; and Melt curve as 95°C for 30 seconds, 47°C for 30 seconds and 95°C for 30 seconds. The qPCR program for HERV-K (HML2) *env* and *gapdh* was as following: 48°C for 30 minutes (reverse transcription step), 95°C for 10 minutes (Holding), 35 cycles of 95°C for 15 seconds, 60°C for 1 minute; and Melt curve as 95°C for 15 seconds, 60°C for 30 seconds and 95°C for 15 seconds. RNA preparations were confirmed to lack detectable DNA or contained very low quantity by performing a quantitative PCR lacking reverse transcriptase in SYBR green system, RT-enzyme mix was replaced by 0.16µl of nuclease-free water. RT-qPCR was conducted using a StepOnePlus™ Real-Time PCR System (Applied Biosystem, Cat # 4376600). Results was analysed by the $2^{-\Delta\Delta Ct}$ method.

For quantification of PcEV *pol* and SIV *gag* and ltr transcripts in plasma, 5µl of DNAse-treated RNA preparation was used per qPCR reaction. The qPCR reaction was performed in one-step using RNA Ultrasense one-step quantitative RT-PCR kit (Invitrogen, Cat # 11732927). For PcEV *pol*, the qPCR reaction consisted of 25.9µl of nuclease-free water, 1µl of each forward and reverse primers (5µM each, final concentration: 100nM), 2.5µl of PcEV-specific Taqman probe (10µM), 10µl of 5X Master mix, 0.1µl of ROX reference dye, and 2.5µl of enzyme mix (superscript III Reverse Transcriptase and Platinum Taq polymerase). The qPCR program was 48°C for 30 minutes reverse transcription step, 95°C for 5 minutes hold step, 40 cycles of 95°C for 15seconds and 61°C for 1minute. For SIV *gag* and ltr, the mix contained 25.9µl of nuclease-free water, 1µl of each forward and reverse primers (5µM each, final concentration: 100nM), 2.5µl of SIV-specific probe (10µM). The qPCR program was 52°C for 60 minutes reverse transcription step, 95°C for 10 minutes hold step, 40 cycles of 95°C for 30 seconds and 61°C for 90 seconds. RT-qPCR was conducted on Mx3005P QPCR System (Agilent Technologies, Cat # 401456). PcEV *pol* and SIV *gag*/ltr in the plasma were represented as a number of copies of PcEV and SIV per ml of plasma.

For quantification of cellular-associated transcripts, each RNA preparation was diluted to 10ng/µl, so that 50ng (5µl) was added in each qPCR reaction. For PcEV *pol* and SIV *gag*/ltr, the qPCR was conducted as stated above. For GAPDH, the mix contained 25.9µl of nuclease-free water, 1µl of each forward and reverse primers (5µM each, final concentration: 100nM), 2.5µl of GAPDH-specific probe (10µM). The qPCR program was 52°C for 60 minutes reverse transcription step, 95°C for 10 minutes hold step, 40 cycles of 95°C for 30 seconds and 61°C for 90 seconds. For STAT1, Sybr green-based qPCR was used. The kit used was RNA-to-Ct Power Sybr green (Applied Biosystem, Cat# 4389986). The mix consisted of 15.6µl of nuclease-free water, 4µl of each forward and reverse primers (10µM each, final concentration: 400nM), 25µl of 2X Master Mix (containing Sybr green and polymerase), and 0.4µl of RT-enzyme mix (for reverse transcription). The qPCR program was as following: 48°C for 30 minutes (reverse transcription step), 95°C for 10 minutes (Holding), 40 cycles of 95°C for 15 seconds, and 60°C for 1 minute; and Melt curve as 95°C for 15 seconds, 60°C for 15 seconds and 95°C for 15 seconds. RNA preparations were confirmed to lack detectable DNA or contained very low quantity by performing a quantitative PCR lacking reverse transcriptase for PcEV, SIV and GAPDH, systematically. In RNA Ultrasense system, enzyme mix was replaced by 0.2µl of Platinum Taq polymerase (10U/µl) (Invitrogen, Cat# 10966), while in Sybr green system, RT-enzyme mix was replaced by 0.4µl of nuclease-free water. RT-qPCR was conducted on Mx3005P QPCR System. PcEV *pol* and SIV *gag*/ltr in tissues were represented as a number of copies of PcEV or SIV per 1000 copies GAPDH

GAPDH was picked as a housekeeping gene as its expression is relatively stable in the same tissues from different individual in human (intra-tissue comparison) (Barber et al., 2005), and as previously used housekeeping gene in the macaque for comparing gene expression in tissues (Ferguson et al., 2014). The comparison of GAPDH expression revealed that it differs significantly in PBMC only (inter-tissue comparison, supplementary information, Fig S5). Also, the variation of expression across the same tissue taken from different individuals (intra-tissue comparison) was estimated by the coefficient of variation (supplementary information, Fig S5).

7.18 *In vitro* transcription.

pBS-PcEV was linearized by Hind III which restriction site is located 16nt downstream PcEV sequence. The reaction mix included 1µg of pBS-PcEV plasmid (0.2µg/µl), 2U/µl final concentration of enzyme and 15µl of water/enzyme buffer. The linearization was carried at 37°C for an hour. *In vitro* transcription was carried using MAXIscript kit (Ambion, Cat # AM1312) following manufacturer's instructions. The reaction was prepared as following: 5µl of linearized pBS-PcEV (0.2µg/µl), 2µl of 10X transcription Buffer, 1µl of each nucleotide solutions (10mM) ATP, GTP, CTP and UTP, 2µl (1.5U/µl final concentration) of T7 enzyme mix (15U/µl stock), and finally 7µl of nuclease-free water. The tube was gently flicked and incubate for 1h at 37°C. The reaction was treated with 1µl of TURBO DNAse I (2U/µl) for 15 min at 37°C to remove DNA. Transcripts were purified as followed: 30µl water was added to the DNase I-treated transcription reaction to bring the volume to 50µl. 5µl of 5 M Ammonium Acetate was added to the mix and vortexed. 3 volumes of 100% ethanol (150µl) was added. The reaction mix was chilled at –20°C for 30 minutes, spun for 20 minutes at maximum speed (14800rpm). Supernatant was discarded and the pellet was washed with 50µl of 70% Ethanol, then spun for 20 minutes at maximum speed again. For complete removal of unincorporated NTPs, the transcripts went through an additional purification step which consisted of adding 5µl of 5M Ammonium Acetate to the mix and vortex. 3 volumes of 100% ethanol (150µl) was added. The reaction mix was chilled at –20°C for 30 minutes, spun for 20 minutes at maximum speed (14800 rpm). The supernatant was discarded and the pellet was washed with 50µl of 70% Ethanol, then spun for 20 minutes at maximum speed again. The pellet was finally resuspended in 20µl of nuclease-free water, followed by nanodrop measurements. The nanodrop indicated 432.5ng/µl with ratios 260/280 = 2.12 and 260/230 = 2.29 for pBS-PcEV transcripts.

7.19 Standard curves.

The PcEV transcripts preparation was DNAse-treated and subjected to 10-fold serial dilutions. The copy number of PcEV for each dilution was estimated as followed: 2µl of the lowest point of the standard curve was further diluted in 2400µl of RT-qPCR mix, corresponding 48 reactions of 50µl. After amplification, the frequency of positive

reactions was determined (f = 8/48) and the copy number per reaction (for 5μl template of each standard) was estimated by the formula: copy number = - ln f (= - ln (8/48) = 1.79), multiplied by 2.5 as frequency obtained correspond to 2μl of the lowest standard. The copy number per reaction was 4.475.

The Ct value obtained for each plasma sample was reported to the standard curve to estimate the PcEV copies per reaction (for 5μl of plasma RNA preparation). The PcEV copies per ml for each plasma sample was obtained taking into account the extraction procedure. As RNA from 140μl or 1ml of plasma was extracted via RNEAsy column, and eluted in 50μl or 100 μl, respectively, the copy number per reaction was multiplied by an enrichment factor of 71.43 (1ml/140ul x 50/5ul) or 20 (1ml/1ml x 100/5), respectively. In further details, the copy number n in 5ul should be multiply by 10 to get the copy number in 50ul (10 x n). The amount of RNA eluted in 50ul was extracted from 140ul, so there are 10 x n copies in 140ul of plasma. So, in 1ml of plasma, there are (1000ul/140ul) x 10 x n copies = 71.43n copies. Similarly, the copy number n in 5ul should be multiply by 20 to get the copy number in 100ul (20 x n). The amount of RNA eluted in 100ul was extracted from 1ml, so there are 20 x n copies in 1ml of plasma. So, in 1ml of plasma, there are (1000ul/1000ul) x 20 x n copies = 20n copies.

Similarly, the lowest standard for SIV possess a copy number of 2.74. As above, the copy number per reaction obtained through standard curve was multiplied by an enrichment factor of 71.43 or 20. SIV standards used were from SIVmac251 RNA plasma series (Berry et al., 2008; Ham et al., 2010).

Standards for GAPDH and STAT1 are kind gifts from Giada Mattiuzzo (NIBSC) (Ferguson et al. 2014). They are made of GAPDH and STAT1 amplicon cloned in a plasmid, separately.

7.20 Immunohistochemistry

Case slides were kindly provided by Derriford hospital Neuropathology department (Dr David Hilton and Dr Phil Edwards). Each slide contains a section from normal nerves or schwannoma. Slides were baked at 60°C for 1 hour. Dewaxing step: The slides were dewaxed by 2 washes in 100% xylene (5 minutes per wash), followed by two washes

in 100% ethanol (5 minutes per wash). The slides were then washed in running tap water for 5 minutes. The samples were then blocked in methanol containing 3% hydrogen peroxide (H_2O_2) for 30 minutes, followed by washing in running tap water for 10 minutes. Pre-treatment to reveal antigens was performed in Tris/EDTA buffer (2.4mg/ml Tris, 0.2mg/ml EDTA, 2mM HCl pH9.0) for HERV-K Gag and in Citrate buffer (2.1mg/ml citric acid, 10mM NaOH, pH6.0) for HERV-K Env, heated in microwave, full power, for 30 minutes. Each antibody was tested on tonsils in different conditions and pre-treatments, the one that reveal a better stain of the antigen per antibody was chosen. After a quick cooling in water, a hydrophobic ring surrounding the tissue was drawn with an ImmEdge pen, to avoid leakage during incubation time with blocking serum or antibodies, before the slides were immersed in 0.05M Tris-HCl buffer (6.05mg/ml Tris, 8mg/ml sodium chloride, 1:1000 Tween 20, pH7.6) for 5 minutes. The slides were laid out on incubation trays with piece of damp tissue beneath the slides for humidity. Normal Horse Serum from Vectastain kit (Vector, Cat # PK-6200) was diluted in 0.05M Tris-HCl buffer (1 drop in 5ml) and was used to cover the tissue section within the ring for 30 minutes. This step blocked non-specific binding sites. Then, liquid on slides was drained onto disposable absorbent tissue, followed by incubation with primary anti-HERV-K (HML2) Gag or Env antibody at a dilution of 1:100, overnight at 4°C. After overnight incubation, the primary antibodies were drained on absorbent tissue and slides were washed 2 times by immersion in 0.05M Tris-HCl buffer for 5 minutes per wash. Then sections were incubated for 30 minutes with secondary antibody solution. The secondary antibody solution is made of normal horse serum and biotinylated universal antibody (Vectastain, 2 drops each in 5ml 0.05M Tris-HCl buffer). Then secondary antibody was washed off by immersion in 0.05M Tris-HCl buffer, 2 times for 5 minutes. Then Streptavidin/Biotin Complex-ABC (Vectastain, 2 drops of both reagent A and reagent B in 5ml of 0.05M Tris-HCl buffer) was applied on slides for 30 minutes. Then the slides were washed two times by immersion in 0.05M Tris-HCl buffer 2 times for 5 minutes per wash. The slides were then laid out on the trays and 3,3′-Diaminobenzidine tetrahydrochloride (DAB) (Sigma, Cat # D5905), dissolved in 5 ml of distilled water and filtered, was applied on the tissue sections for 5 minutes. Excess was taped off onto an absorbent tissue and slides were washed in running tap water for 10 minutes. Then when necessary, the slides were immersed in DAB enhancer solution (4g of copper sulphate and 7.2g sodium chloride in 1l of distilled water) for 2 minutes to increase the signal when needed, followed by

a 5-minutes wash in running water. The tissue sections were then counterstained using haematoxylin for 1-2 minutes followed by a 10-minute wash in running water. Next, the slides were quickly washed 2 times in 100% ethanol and 2 times in 100% xylene, then a coverslip was mounted on top of the section on the slide using DPX mounting solution. Staining is presented as brown and intensity was scored unblinded by a third-party neuropathologist, David Hilton, NHS, Plymouth Hospital. Results are presented as: -, no staining; -/+, equivocal; +, positive; ++, strong positive; +++, very strong positive.

7.21 Immunofluorescence

Cells were seeded at a concentration of 3000 cells/well on a Lab-Tek, and grown until they were confluent. After drug treatment, cells were washed with PBS, fix with 4% PFA in PBS for 10 minutes at room temperature. Cells were washed with PBS again and permeabilised with 0.2% Triton X100 in PBS for 60 minutes. After permeabilisation, cells were washed three times with PBS for 5 minutes each. Cells were blocked using 10% serum 1% BSA in PBS for an hour at room temperature. Then blocking solution was washed with PBS for 5 minutes at room temperature. Cells were incubated in a primary antibody solution made by diluting antibody in 1% BSA in PBS overnight, the primary antibodies were against: Ki67, a maker of cell proliferation, used at 1:100; c-Jun, a protein associated to proliferation and demyelination, used at 1:100; HERV-K Env at 1:50 and HERV-K Gag at 1:50. Secondary fluorescent antibody (section 7.3) were prepared in 1% BSA in PBS at a dilution of 1:500 and applied for an hour at room temperature and sheltered from light. Cells were then washed three times in PBS for 5 minutes each. DAPI nuclear staining was applied at a concentration of 1:500 in PBS and incubated for 10-15 minutes. Cells were then washed three times for 5 minutes each. The Lab-Tek chambers were removed and Vectashield was used to apply a coverslip on top of the cells. The edges of coverslip were sealed using clear nail polish and the slide was stored at 4°C in the dark until microscopy. Confocal Microscopy was performed by Dr Sylwia Ammoun.

7.22 Exosome isolation

Whilst the cells were still in culture, the standard media was removed and replaced with media made with exosome-depleted FBS (System Biosciences, Cat # EXO-FBS-250A-

1), as preliminary trials indicated that the exosomes present in standard FBS were affecting our results. The cells were cultured in the exosome-depleted medium for 72hours at which point the media was removed for exosome isolation. The exosomes were extracted using Total Exosome Isolation Reagent (Thermofisher, Cat # 4478359) and Exoquick Exosome Isolation (System Biosciences, Cat # EXOQ5A-1) according to the manufacturer's instructions. Briefly, exosome-depleted medium from cultures was collected in a 15ml or 50ml tube (depending on volume of medium available in culture) and centrifuged at 3000 x g for 15 minutes to remove cellular debris. The supernatant was transferred to in a separate tube, 63µl of ExoQuick Exosome Precipitation Solution (ExoQuick^{TM}, Cat # EXOQ_A-1) was added for each 250µl of culture medium, mixed by flicking and refrigerate overnight at 4°C. Then, the mixture was centrifuged at 1500 x g for 30 minutes. The supernatant was discarded (or kept for exosome-free control) and the pellet was further centrifuged at 1500 x g for 5 minutes to remove residual fluids. The pellet containing exosomes was resuspended in RIPA lysis buffer and stored at -20°C until required.

Total Exosome Isolation Reagent was also used to pellet exosomes from schwannoma's culture medium, according to the manufacturer's instructions. The method was similar and no specific differences in exosome fractions obtained among the project, was observed.

CD63 is a tetraspanin present on the membrane of exosomes and intracellular vesicles. CD63 was included in the western blotting panel as a positive control for exosome isolation (Thery et al., 2006).

Exosomes isolations were mainly performed by Shona Reeves, Undergraduate student from University of Bath, supervised by the thesis author.

7.23 FACS

FACS was used to determine the proportion of transfected cells (GFP+). Transfected cells were trypsinised and washed 3 times in buffer made of 1% FBS in PBS. Then cells were resuspended in 300µl of the same buffer. Non-transfected cells were used as control. Cells suspensions were analysed on FACS Accuri C6 flow cytometer (BD Biosciences)

7.24 HERV-K (HML2) LTR mapping

Alignments were build using MEGA6 software. LTRs from the following full length HERV-K (HML2) loci sequences were aligned: HERV-K113 (19p12b), HERV-K116 (1p13.1), HERV-K102 (1q22), HERV-K106 (3q13.2), HERV-K117 (3q27.2), HERV-K107 (5q33.3), HERV-K109 (6q14.1), HERV-K108 (7p22.1), HERV-K118 (11q22.1), HERV-K119 (12q14.1), HERV-K 12q13, HERV-K103 (10p22.1), HERV-K101 (22q11.21), ERVK4 (3q21.2), HERV-K115 (8p23.1a), HERV-K60 (21q21.1), HERV-K104 (5p13.3), ERVK19 (19q11), HERV-K 2q21.1. HERV-K (HML2) was annotated manually based on previous studies: only 7 TFs motifs have been tested experimentally: OCT4 (Grow et al., 2015), NF-κB and NF-AT (Gonzalez-Hernandez et al., 2012), MITF-M (Katoh et al., 2011), Sp1 and Sp3 (Fuchs et al., 2011), YY1 (Knössl et al., 1999). TEAD binding motif TGGAAT from (Vassilev et al., 2001) binding motifs was searched and annotated manually.

7.25 Overview of the macaque genome.

A more detailed analysis of data presented in Magiorkinis et al. (2015) suggested that PcEV is one of the 3 recently active ERV lineages in the macaque, with two other lineages (SERV, CERV) that have shown similar levels of copying over the last few million years (Fig. S2). The relative youth of these lineages among the entire complement of macaque ERVs is shown in Fig. S3. This type of analysis also shows the macaque to be intermediate between the mouse and human genomes in the number of recently integrated ERV loci (Fig. S4).

7.26 Extraction and alignment of macaque ERV locus sequences

Bioinformatic analyses used UCSC's Genome Browser website and its BLAT tool (Kent, 2002). We worked primarily with the most recent and reportedly best reference genome sequence of the rhesus macaque, rheMac8 (Mmul_8.0.1). This assembly derives from Zimin et al.'s (2014) (short-read) Illumina whole genome sequencing (MacaM) with later closing of some gaps using (long-read) PacBio sequencing of the same individual animal (comment in GenBank accession GCA_000772875.3)

We compared the sequences of loci with those from an earlier build of the rhesus macaque genome, rheMac2 (Richard A. Gibbs, 2007) (GenBank accession GCA_000002255.2). This is an earlier (short-read) whole genome sequencing project but it is largely from the same (female) individual (ID 17573) plus some sequencing of an unrelated male used in finishing and in providing a Y chromosome. We also compared the sequences to those in another species of macaque, the cynomolgus (crab-eating or long-tailed) macaque. These two species are very closely related and so have a very similar genome sequences: they are estimated to share a common ancestor only 0.91 ± 0.11 million years ago, which is more recent than the common ancestor of all rhesus macaques (Osada et al., 2008). The reference genome sequence of the cynomolgus macaque, macFas5, appears to have been built independently of the rhesus genome using a range of short-read technologies (comment in GenGank accession GCA_000364345.1). The other rhesus build available on the UCSC website, rheMac3, is very fragmentary.

PcEV was described initially from the baboon (Mang et al., 1999) (with GenBank reference AF142988) and was erroneously referred to as BaEV in (Magiorkinis et al. (2015) (see SI). We initially used this GenBank sequence to find and download the sequences of several PcEV loci in the macaque using the UCSC Genome Browser. A multiple alignment was then made of the downloaded sequences using MEGA (Tamura et al., 2013, 2011) and a new consensus reference sequence built that had full-length ORFs (Open Reading Frames) in all genes (sequences from different loci differed only by a few percent so alignments were unambiguous). This macaque reference was then used to re-search the macaque genome as above. Multiple alignments were then made manually from downloaded sequences, and the individual loci examined visually for full-length ORFs and other motifs.

7.27 Dating integrations

We dated integrations using the nucleotide divergence between the LTRs of the proviruses. These LTRs form the flanks of a provirus (the complete integrated DNA form of a retrovirus) and are identical at the time of integration, accumulating substitutions at the host background rate.

We used the following equation to calculate locus age:

Estimated age = (a/b)/r*2 where

a = number of mismatches between the two LTRs

b = length of the LTR

r = estimated rate of nucleotide substitution. We need to in effect divide the final age by two to find the integration date because substitutions will have occurred along the two branches leading to the 5' and the 3' LTRs (alternatively, we could multiple the LTR length by two to reflect this). Our rate of nucleotide substitution is taken to be $1.0x10^{-9}$/nucleotide/year based on analyses of primate ERV sequences (Magiorkinis et al., 2015). This value is about half the rate estimated for mammals generally but close to estimates for the rate of neutral molecular evolution in Old Word Monkeys and Apes (Subramanian and Kumar, 2003; Yi et al., 2002) and was also used by Osada et al. (2008) to date the divergence of the two macaque species.

7.28 Prediction of Transcription factor binding sites on PcEV LTRs

Transcription factor binding motifs was predicted with ALGGEN-PROMO online tool (Farré et al., 2003; Messeguer et al., 2002) (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3). The transcription factors and binding sites were select for primates only. The following primate-specific transcription factor binding sites were searched for ([Accession number in TRANSFAC public database]): AP-1 [T00029]; NFI/CTF [T00094]; C/EBPalpha [T00105]; NF-Y [T00150]; CREB [T00163]; c-Rel [T00168]; CTF [T00174]; GATA-1 [T00306]; GATA-2 [T00308]; GATA-3 [T00311]; GR-alpha [T00337]; NF-1 [T00539]; NF-AT1 [T00550]; C/EBPbeta [T00581]; C/EBPdelta [T00583]; NF-kappaB [T00590]; NF-kappaB1 [T00593]; RelA [T00594]; POU2F1 [T00641]; Sp1 [T00759]; TBP [T00794]; YY1 [T00915]; STAT1alpha [T01492]; STAT3 [T01493]; STAT1beta [T01573]; NF-YA [T01804]; GR-beta [T01920]; NF-AT2 [T01945]; NF-AT1 [T01948]; STAT1 [T04759]; GR [T05076]. The dissimilarity between the query and the binding site matrixes from the database was fixed at a threshold of 5%. The sequence submitted as a query was PcEV LTR consensus. It was obtained as a sequence made with the most represented nucleotide for each position from an alignment built from LTRs of all PcEV loci gathered. The alignment includes PcEV LTR sequences from NC_022517.1 (Kato et al., 1987) and AF142988.1 (Mang et al., 1999).

The list of transcription factors to predicted was based on motifs described in human HERV-W LTRs and experimentally tested CCAAT box, TATA box, POU2F1 (Oct1) (Lee, 2003); on previous *in silico* analysis of PcEV LTR (GATA-1/-2, CCAAT box, Sp1) (Mang et al., 1999); Enhancer sequences in MuLV (NF1, GRs) (Speck and Baltimore, 1987); transcription factors suggested to be involved in inflammation-stimulated retroviruses (NF-κB/NF-AT1-induced HERV-K(HML2) (Gonzalez-Hernandez et al., 2014) and Tax-induced HERV-W (Toufaily et al., 2011); NF-κB-induced HERV-W (Uleri et al., 2014); STAT3 activation and CREB recruitment in HIV infection (Marzio et al., 1998)). The sequence submitted as a query was PcEV LTR consensus, first sequence in Figure 31. The alignment includes PcEV LTR sequences from NC_022517.1 (Kato et al., 1987) and AF142988.1 (Mang et al., 1999). The dissimilarity threshold was fixed at 5%. They could be a bias as most of the TF motifs from primates are from human. The database is poor on motifs from other primate species, as there was not any selection for binding sites corresponding to *haplorrhini* and *catarrhini* for example. This is likely due to the lack of experimental test for transcription factors in species from those groups. However, the macaque and human genome are almost 97% similar, so we suspect that transcription factor binding sites are likely conserved.

Transcription factor binding sites based on matrix provided on ALGGEN-PROMO online tool: *Core sequence is underlined.*

TBP [T00794] – RRKRTATAAA

GR-beta [T01920] – AATKD

GR-alpha [T00337] – WNAGG

C/EBPbeta [T00581] – TTGN

NF-1 [T00539] – TTGGSMYR

C/EBPalpha [T00105] – DATTGNB

STAT3 [T01493] – YKVYTTCCCSGMAV

GATA-2 [T00308] – DCHYTATCD

GATA-1 [T00306] – TATCNN

NF-Y [T00150] – ATTGGYYH

YY1 [T00915] – CCAT

7.29 Statistical Analysis

For the HERV-K (HML2) study in sections 8.1, 8.2 and 8.3, student's two-tailed t-tests and one-way ANOVA with post-hoc Tukey honestly significant difference (HSD) test were used for statistical analysis of the data. On graphs, ns represents not significant, a p-value < 0.05 was represented as *; a p-value < 0.01 as **; and a p-value < 0.001 as ***. In the graphs, mean +/- standard error of the mean are given.

For the PcEV study in section 8.4, Graphpad Prism version 6.01 was used for statistical analysis: Mann-Whitney test was carried for comparing non-paired samples in SIV- vs SIV+ groups; Kruskal-Walis test was used to compare PcEV levels in tissues altogether; Shapiro-Wilk test for normality distribution of cell associated STAT1, SIV, PcEV and plasma SIV values; since most data set presented distributions of values that were significantly different from normal according to Shapiro-Wilk test, we present the correlation coefficients r and p-values from non-parametric Spearman test; where logarithmic transformation (base 10) of values were not significantly different from normal, we also present r and p-value from parametric Pearson test; Fisher's exact test to assess whether PcEV released in the plasma is commoner in infected compared to animals; Wilcoxon matched-pairs signed rank test to compare cellular and plasma expressions of SIV to cellular PcEV and STAT1.

8 RESULTS

8.1 HERV-K (HML2) protein up-regulation in schwannomas and potential therapies

8.1.1 Background

Schwannomas are tumours from the peripheral nerve system, developed from Schwann cells. The loss of Merlin, a tumour suppressor, is observed in 100% of neurofibromatosis type-2-related schwannomas (Evans, 2000; Hanemann and Evans, 2006) and approximately 70% of sporadic schwannomas (Agnihotri et al., 2016). Merlin deficiency is the key event which likely drives tumour formation (Evans, 2000; Hanemann, 2008; Hanemann and Evans, 2006; Hilton and Hanemann, 2014). In fact, Merlin loss in schwannomas induces overexpression of many receptors involved in cell proliferation and survival such as PDGFR (platelet-derived growth factor receptor), IGF-1R (insulin-like growth factor receptor), ErbB2 (epidermal growth factor receptor 2), VEGFR (vascular endothelial growth factor receptor) and Axl (Hilton and Hanemann, 2014). Interestingly, re-introducing Merlin into schwannoma primary cells decreases tumours' features as high proliferative level, survival and cell death inhibition (Ammoun et al., 2014; Schulze et al., 2002).

To date, therapies for schwannomas include surgery and radiosurgery. Such methods can be limited when the tumours are present in multiple sites in the body. Also, surgery presents risks because of the invasiveness that can cause further damage, and radiosurgery can induce undesirable effects such as additional mutations. Unfortunately, current chemotherapy is not effective for those tumours, stressing the need for new therapeutics.

HERV-K (HML2) expression is considered to be a common feature in cancers including melanoma, germ cell tumours, seminomas, breast cancer, ovarian carcinomas, prostate cancer and lymphoma (Hohn et al., 2013). For that reason, HERV-K (HML2) proteins are interesting tumour associated antigens. In fact, monoclonal antibody binding to HERV-K (HML2) Env has been suggested to reduce proliferation (Wang-Johanning et al., 2012).

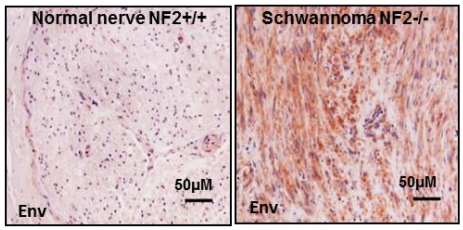
In the present section, measuring HERV-K (HML2) protein expression in these tumours is part of the objectives as it has never been assayed before in this tumour. Also, I explored the possibilities of HERV-K (HML2) proteins to be targeted as an antigen for monoclonal antibody therapy, and the potential of an antiretroviral drug, ritonavir, to be used as treatment for schwannomas.

8.1.2 HERV-K (HML2) expression in schwannomas and normal nerves

Immunohistochemistry was used to find tissues that harbor HERV-K (HML2) Env and Gag antigens among tumours and normal nerves. Ten and 15 cases were assayed for the presence of HERV-K (HML2) Env and Gag, respectively. The general picture was that staining in normal nerves were weaker than schwannomas (Figure 11A and B), suggesting a general overexpression of HERV-K (HML2) proteins in tumour. Regarding the presence of Env in tumours, 8/10 (80%) cases were positive with varying intensities, from weakly to very strongly positive. 1 in 10 (10%) was equivocal (hard to state whether it is positive or negative), and 1/10 (10%) was negative. In normal nerves, 3/10 (30%) cases were positive. 2/10 (20%) were equivocal, and 5/10 (50%) were negative. Env is also expressed in the normal nerves, but to a lesser frequency than in schwannomas (table in Figure 11A).

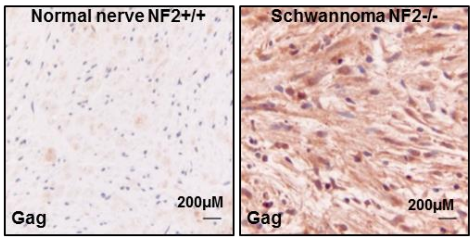
Regarding HERV-K (HML2) Gag in schwannomas, 13/15 (87%) were positive, and 2/10 (13%) were equivocal. In normal nerves, 9/15 (60%) were negative, 3/15 (20%) were equivocal, and 3/15 (20%) were positive. The frequency of schwannomas expressing Gag was higher than normal nerves. Also, the intensity of the staining was generally higher in schwannoma tissues (Figure 11B). Interestingly, there were no negative schwannomas, suggesting Gag is found more often than Env in tumours (table in Figure 11B).

A



Normal nerve	Schwannoma
+	+
+	+/-
+/-	+++
+/-	++
+	++
-	+++
-	+
-	-
-	+
-	+++

B



Normal nerve	Schwannoma
+	++
+	++
+/-	++
+/-	++
+	+
-	+
-	+
-	+
-	+
-	+
-	+
-	+
-	+
-	+/-
+/-	+/-

Figure 11 : Immunohistochemistry analysis of biopsies from schwannomas and normal nerves.

Sections of paraffin-embedded tissues were incubated overnight with (A) an anti-HERV-K (HML2) Env or (B) an anti-HERV-K (HML2) Gag after pre-treatment in heated Citrate or EDTA buffer for 30 minutes, respectively; followed by biotin-conjugated universal secondary antibody (Vectastain) and streptavidin-biotin complex ABC (Vectastain). The immunostaining was revealed with 3,3′-Diaminobenzidine tetrahydrochloride. Staining of HERV-K (HML2) Env and Gag was denser in biopsies from schwannomas than from normal nerves. The frequency of individual expressing HERV-K (HML2) Env and Gag is higher in biopsies from patients (Env: 9/10; Gag: 15/15) than in healthy controls (Env: 5/10; Gag: 6/15).

8.1.3 HERV-K (HML2) expression in primary cultures of schwannomas and Schwann cells

Human primary schwannoma and Schwann cells can be isolated and cultured *ex vivo*. It provides a model for testing the effects of a drug on the proliferation of tumour cells. Cells from cultures were assessed for HERV-K (HML2) expression using western blotting (n=4) and immunofluorescence (n=3). After cell lysis and separation by electrophoresis, proteins were transferred onto PVDF membranes and probed with HERV-K (HML2) Env and HERV-K (HML2) capsid-specific antibodies, respectively. MCF-7 cell lysate was used as a positive control for Env detection, as they have been previously shown to express HERV-K (HML2) Env (Wang-Johanning et al., 2012). The blots were imaged and densities of the bands obtained from antibodies binding to the target were quantified. The *in silico* analysis of HERV-K (HML2) Env sequence and a previous report (Kämmerer et al., 2011) suggested that the commercial antibody used reacts against the transmembrane-unit (TM, 27 kDa un-glycosylated). Env is produced as a poly-protein, formed by a transmembrane and a surface unit (SU, 60 kDa un-glycosylated), that is cleaved by cellular furin or furin-like proteases (Moulard and Decroly, 2000), they form a heterodimer at surface by protein-protein interaction. Env possesses a total of 11 N-glycosylation sites, 4 on the TM and 7 on the SU. The size of an N-glycosylation in mammal cells is between 3-4 kDa (estimated by the addition of the molecular weight of glycans that constitute a group of N-glycosylation) (Varki et al., 1999). The TM should be 40 kDa and 36 kDa when 4 and 3 sites are glycosylated, respectively. Interestingly, Kämmerer et al. (2011) reported that the antibody recognise a 55 kDa product, this was suggested to be an overglycosylated TM by the corresponding author (J. Denner, personal communication, 02/2017). The Env-specific antibody used cannot detected SU alone. SU and TM together (SU+TM) should be detectable at a size of 70 kDa un-glycosylated. When fully glycosylated, Env should be of a size slightly above 100 kDa, 103-114 kDa. The highest band described for the full length Env (SU+TM) is approximately 90 kDa (Henzy and Coffin, 2013; Tönjes et al., 1997). That corresponds to 5-8 N-glycosylations (3-4 kDa per group of N-glycosylation added). Env blots in MCF-7 lysate exhibited a set of 5 bands mainly: 98 (that would harbor 7-9 N-glycosylations); 70; 55; 40; 38 kDa. The same pattern was generally observed for primary schwannomas (Figure 12A) and Schwann cells, with the 70 kDa band usually being weakly dense. Schwannomas and Schwann cells in culture express Env as a 7-9 N-glycosylated SU+TM, a un-glycosylated SU+TM, a full-glycosylated

(4 times) TM and a 3 N-glycosylated TM. Those are the bands that were quantitated. Density analysis revealed a higher mean intensity for each band in schwannomas than in Schwann cells (Figure 12C). Immunofluorescence applied to Schwann and schwannoma cells confirmed the result above. HERV-K (HML2) Env-specific antibody exhibit a stronger fluorescence in schwannoma than in Schwann cells in culture (Figure 12E), suggesting a higher expression of HERV-K (HML2) Env in the primary schwannomas.

HERV-K (HML2) capsid is produced as part of the Gag polyprotein, which is sequentially made of: Matrix-p15-Capsid-Nucelocapsid (MA-p15-CA-NC). It is emphasized that the bands observed using an anti-capsid antibody are theoretically CA (27 kDa), p15-CA (36 kDa), CA-NC (44 kDa), MA-p15-CA (60 kDa), and Gag full length MA-p15-CA-NC (74 kDa). Schwannoma blots revealed 5 bands mainly (Figure 12B). The bands corresponded to the sizes of MA-p15-CA-NC, MA-p15-CA, p15-CA, CA-NC, CA. Mean intensity for each band was higher in schwannomas than in Schwann cells tested (Figure 12D). Along with this, an immunofluorescence assay performed using HERV-K (HML2) Gag-specific antibody resulted in a stronger labelling in schwannomas than in Schwann cells (Figure 12F). Altogether, our data suggested that primary human schwannoma cells express more HERV-K (HML2) proteins than primary human Schwann cells.

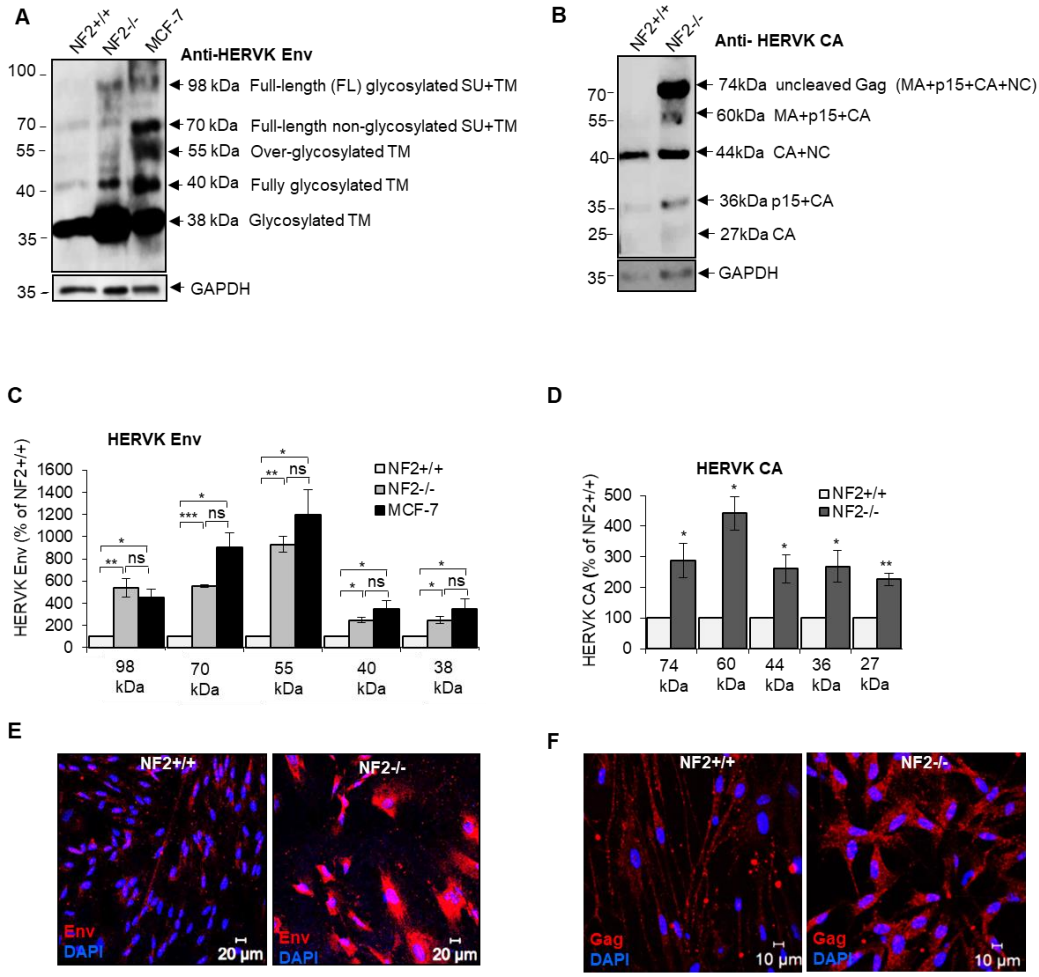


Figure 12 : Primary cultures of schwannoma (NF2-/-) expresses more HERV-K Env and Gag than Schwann cells (NF2+/+).

Tissues from patients and nerves from donors were digested and schwannoma cells (NF2-/-) and Schwann cells (NF2+/+) were put into culture. At convenient confluence (>80%), cells were lysed in RIPA buffer and centrifuged at 13 000rpm for 15 minutes to remove cellular debris. Protein lysates were run on separating gels and transferred onto PVDF membranes. Membranes were incubated overnight (A) and anti-HERV-K (HML2) Env or (B) an anti-HERV-K (HML2) CA. GAPDH was used as loading control for western blot. (A) and (B) show western blot of a representative experiment. (C) and (D) represent mean intensities per HERV-K (HML2) Env and Gag band normalized to GAPDH band intensity, respectively, as a percentage of normalized band intensities observed for NF2+/+ cell lysates. Means are from 4 independent experiments. (E) and (F): Immunofluorescence analysis of schwannoma cells (NF2-/-) and Schwann cells (NF2+/+). Both were seeded in Lab-Tek chambers and culture to a convenient confluence (>80%). Then they were fixed in 4% PFA, blocked, and incubated with anti-HERV-K (HML2) Env or Gag antibodies. A fluorescent-conjugated secondary anti-mouse antibody was used to reveal staining. DAPI was used to mark cell nuclei. The figures (E) and (F) are representative of 3 independent experiments each.

8.1.4 HERV-K (HML2)-specific antibodies reduced proliferation in schwannoma cells but not in Schwann cells

Commercially available anti-HERV-K (HML2) Env and capsid were tested in primary cultures of schwannomas and Schwann cells; a normal mouse IgG was used as a control for "background" effect of antibody treatment. A 3-day incubation with both HERV-K (HML2) specific antibodies resulted in decreased proliferation of schwannoma cells and not in Schwann cells in culture, as measured by ki67 immunofluorescent staining (Figure 13). The experiment was done with Schwann cells (Figure 13B and D) and schwannoma cells (Figure 13A and C) from separate donors (n=3 for Schwann cells; n=4 for schwannomas). The anti-tumour effect of HERV-K (HML2)-specific antibodies seemed specific as incubation with normal mouse IgG control had no significant effect on ki67 staining.

Merlin loss results in dysregulation of cellular pathways in schwannomas. Features are increases in proliferative and survival signals, as monitored by activation and levels of p-ERK, p-FAK, p-AKT using antibodies specific to the phosphorylated form of the proteins, and Cyclin D1 (Ammoun et al., 2008; Ammoun and Hanemann, 2011; Hilton and Hanemann, 2014). The level of those molecules was checked by western blot for primary cultures of schwannoma cells incubated with HERV-K (HML2) Env and capsid-specific antibodies; control cultures were incubated with control mouse IgG. Also, the levels of ERK, FAK, and AKT in total, were monitored using antibodies that recognise both phosphorylated and unphosphorylated forms of these proteins, for checking whether the antibodies added in culture impact the overall levels of each protein. It means that an activation of a pathway is considered when the amount of phospho-proteins increases while the level of total proteins remains stable. In these conditions, we recorded a significant decrease in p-ERK (n=10) and p-AKT (n=9) for cultures incubated with HERV-K (HML2)-specific antibody, in comparison to cultures with control mouse IgG (Figure 14A, B, C and D). The decrease was about 2 times for p-ERK and less than 2 times (~1.5-1.8 times) for p-AKT. Control mouse IgG by itself, when introduced in culture, did not significantly affect the levels of p-ERK and p-AKT. All the antibodies did not affect significantly the levels of cellular total ERK and AKT. On the contrary, control mouse IgG decreased the levels of p-FAK. In comparison to control mouse IgG condition, HERV-K (HML2) Env-specific antibody did not significantly affect p-FAK. Only HERV-K (HML2) capsid-specific antibody further

decreased p-FAK levels. All antibodies did not significantly affect cellular levels of total FAK (Figure 14E and F, n=3). The general picture is then a decrease of pathways involved in proliferation (ERK pathway) and survival (AKT pathway). In concordance to that, Cyclin D1 level was significantly decreased with HERV-K (HML2)-specific antibody in comparison to control mouse IgG (Figure 14G and H, n=4). Though, control mouse IgG reduced significantly Cyclin D1. Our data strongly suggested that HERV-K (HML2)-specific antibodies can reduce schwannoma proliferation. It correlates with data by Wang-Johanning in a breast cancer model (Wang-Johanning et al., 2012).

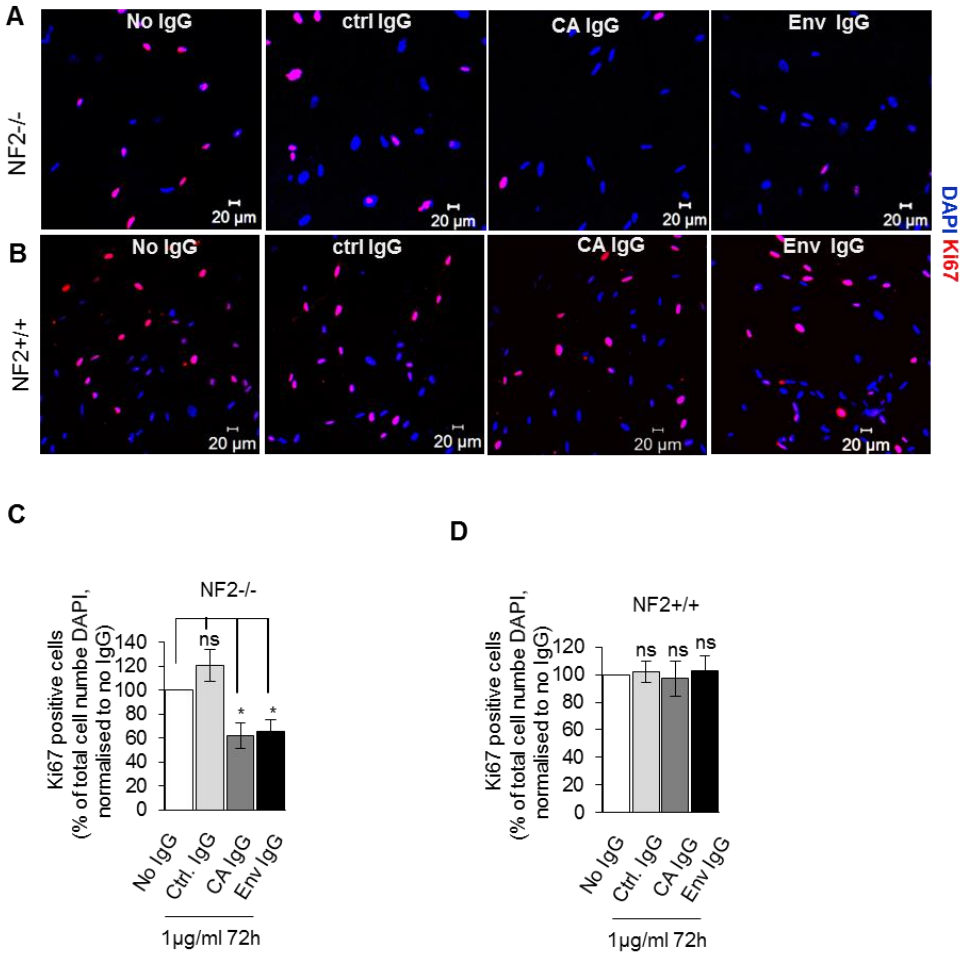
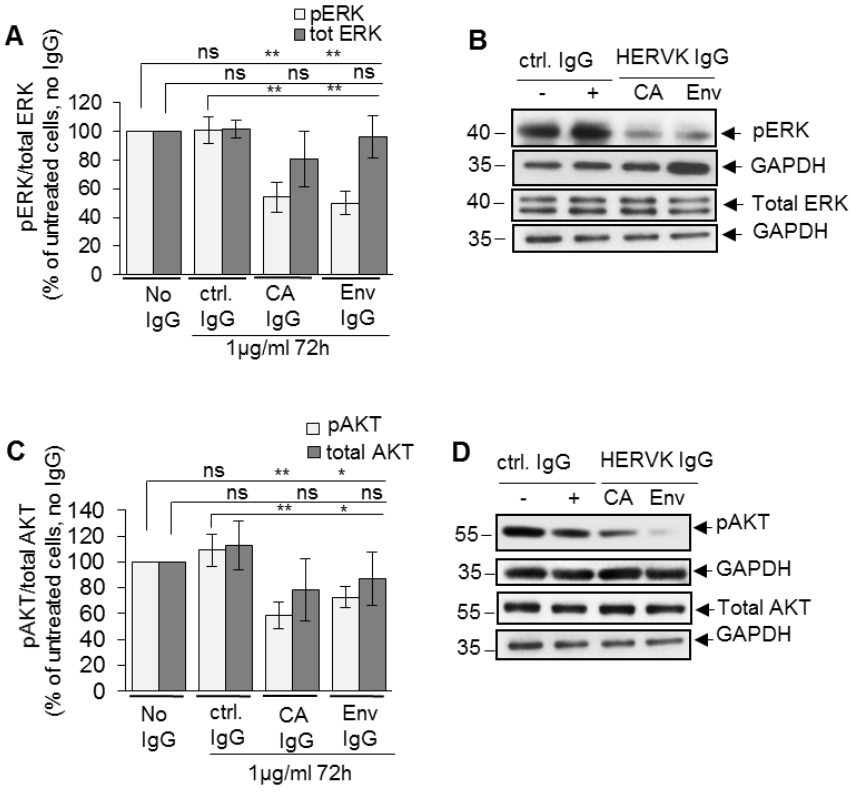


Figure 13 : Effect of anti-HERV-K (HML2) Env and Gag antibodies on the proliferation of schwannoma cells (NF2-/-) and Schwann cells (NF2+/+).

Effect of anti-HERV-K (HML2) Env and Gag antibodies on the proliferation of schwannoma cells (NF2-/-) and Schwann cells (NF2+/+). (A) and (B) show immunofluorescence analysis of a representative experiment. Both cell types were seeded in Lab-Tek chambers and culture to a convenient confluency. Then they were incubated for 72 hours with anti-HERV-K (HML2) Env and Gag at a concentration of 1µg/ml. Normal mouse IgG was used as a control for "background" effect of antibody

treatment. Untreated cells were used as control for all antibody effects. After incubation, cells were fixed in 4% PFA, blocked, and incubated with anti-Ki67 antibody. A fluorescent-conjugated secondary anti-mouse antibody was used to reveal staining. DAPI was used to mark cell nuclei. (C) and (D) represent the percentage of Ki67+ cells per total counted cells (DAPI), normalized to the condition without any treatment. Means of Ki67+ cells for each condition (IgG, anti-HERV-K (HML2) Env, anti-HERV-K (HML2) Gag) were obtained from n=4 and n=3 independent experiments, for schwannomas and Schwann cells, respectively.



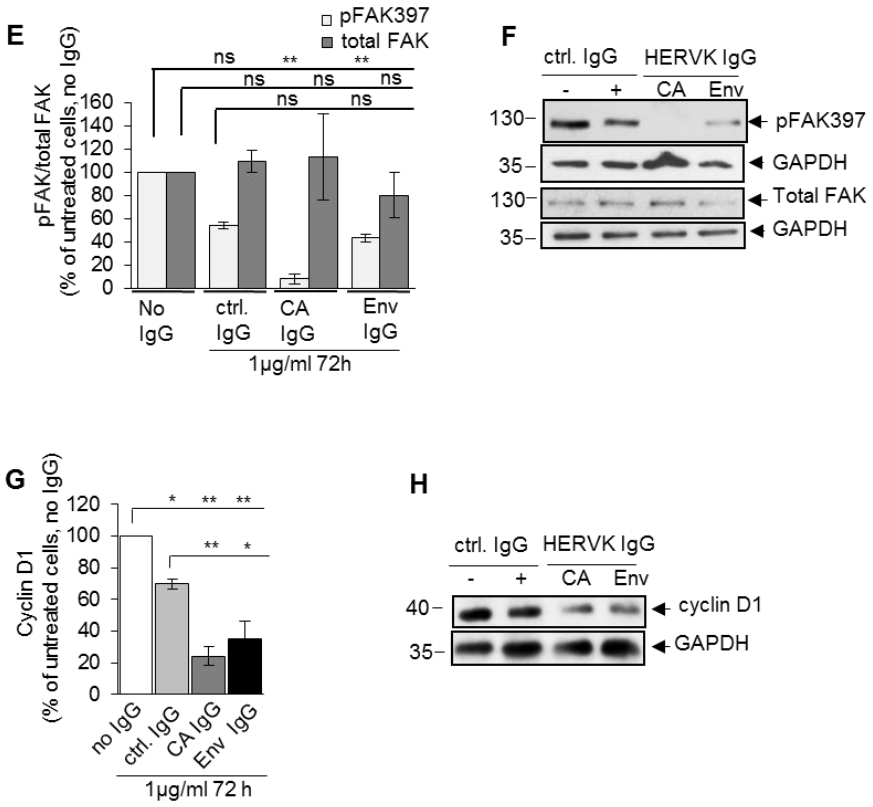


Figure 14 : Effect of anti-HERV-K Env and Gag antibodies on key pathways of schwannoma cells (NF2-/-).

Schwannoma cells were seeded onto 6-well plate. The cells were incubated for 72 hours with anti-HERV-K (HML2) Env or Gag at a concentration of 1µg/ml. Normal mouse IgG was used as a control for "background" effect of antibody treatment. Untreated cells were used as control for all antibody effects. After 72 hours, cells were lysed in RIPA buffer and centrifuged at 13 000rpm for 15 minutes to remove cellular debris. Protein lysates were run on separating gels and transferred onto PVDF membranes. Membranes were incubated overnight with (A-B) an anti-pERK, (C-D) an anti-pAKT, (E-F) an anti-pFAK^{Y397}, or (G-H) an anti-Cyclin D1. GAPDH was used as loading control for western blot. (A, C, E, G) show western blot analysis of a representative experiment. (B, D, F, H) represent mean band intensities normalized to GAPDH band intensity, as a percentage of normalized band intensities observed for untreated NF2-/- cell lysates. Means are from n=10, n=9, n=3, n=4 independent experiments, for p-ERK, p-AKT, p-FAK and Cyclin D1, respectively, using different cell batches from a minimum of four different patients.

8.1.5 Ritonavir reduced proliferation in schwannoma cells but not in Schwann cells

Ritonavir is an antiretroviral drug, used for the treatment of HIV-seropositive individuals. It inhibits HIV protease which cleaves Gag into MA, CA, and NC. Without such cleavage, HIV viral particles released are immature and not infectious (Kohl et al.,

1988). Ritonavir was also shown to inhibit HERV-K (HML2) protease activity, but with a lower efficiency as the dose required to block HERV-K (HML2) protease is more than 6000 times higher than the one that inhibit HIV protease (Kuhelj et al., 2001). In general, the blockade of a retroviral protease in immunoblots experiments shows a decrease in the intensity of products lighter than the full length 74 kDa–Gag, with an increase of the latter. Ritonavir tested on schwannoma cells for 24 hours at 1μM induces a low decrease of the band intensity of the 38 kDa–CA+NC, along with a slight increase of 74 kDa–Gag (Figure 15A and B, n=4). This confirm that HERV-K (HML2) protease is mildly affected by ritonavir. At 1μM, the measured inhibition is incomplete and weak. When added in primary cultures of schwannomas for 72 hours, ritonavir decreased the levels of p-FAK (n=4), p-AKT (n=2), p-ERK (n=3) and Cyclin D1 (n=2) as measured by mean intensity of the blots (Figure 15C and D). Furthermore, ritonavir was tested at 0.1 and 1μM for 72 hours in primary cultures of schwannoma cells from different donors. Untreated cultures were used as controls. Ki67 monitoring revealed a decrease in signal by 30% at 0.1μM already, and by 50% at 1μM (Figure 15E and G, n=4). Interestingly, the same experiment conducted in primary cultures of Schwann cells showed no significant difference in Ki67 levels for both concentration compared to untreated condition (Figure 15F and H, n=3). The decrease of proliferation cannot be attributable to HERV-K (HML2) blockade by ritonavir. All together our findings strongly support the use of ritonavir as a potential drug for schwannomas.

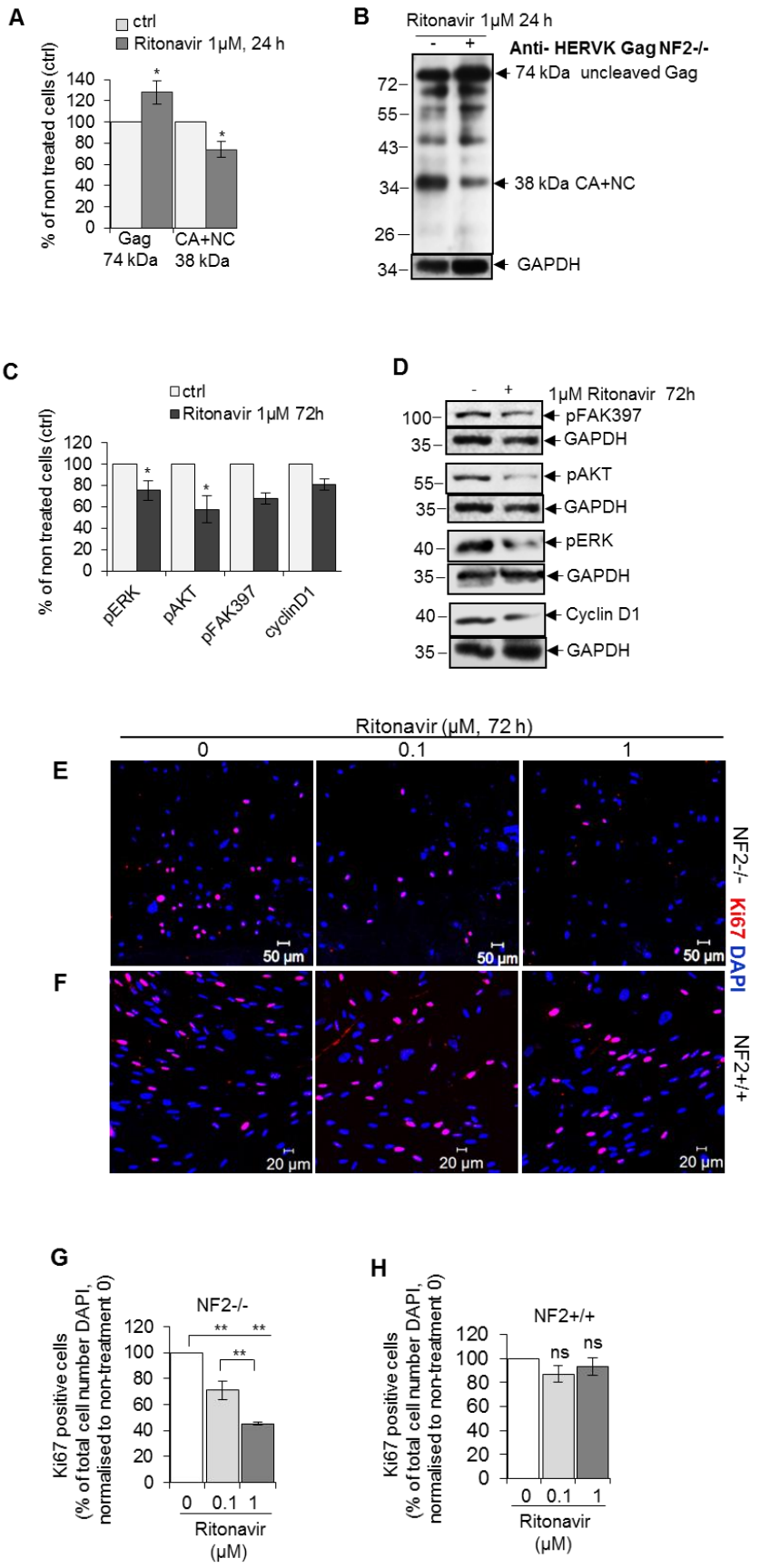


Figure 15 : Effect of Ritonavir on HERV-K (HML2) Gag cleavage, key pathways and proliferation of schwannoma cells (NF2-/-).

Schwannoma cells were seeded onto 6-well plate. The cells were incubated for 24-72 hours with ritonavir at a concentration of 1μM. Untreated cells were used as control. After treatment period, cells were lysed in RIPA buffer and centrifuged at 13 000rpm for 15 minutes to remove cellular debris. Protein lysates were run on separating gels and transferred on PVDF membranes. Membranes were incubated overnight with (A-B) an anti-HERV-K (HML2) Gag to assess ritonavir effect after 24 hours on Gag cleavage. GAPDH was used as loading control for western blot. (B) is a western blot analysis of a representative experiment. Ritonavir at 1μM increase 74 kDa uncleaved Gag, while decreasing 36 kDa cleaved Gag (p15+CA). (A) Represents means of n=4 independent experiments. Membranes were incubated overnight with (C-D) an anti-pERK, an anti-pAKT, an anti-pFAK^{Y397}, or an anti-Cyclin D1. GAPDH was used as loading control for western blot. (D) shows blots of representative experiments for each pathway. Ritonavir at 1μM for 72 hours decreased pERK, pAKT, pFAK^{Y397} and cyclin D1 in comparison to untreated cells. (C) Represents mean band intensities normalized to GAPDH band intensity, as a percentage of normalized band intensities observed for untreated schwannoma cell (NF2-/-) lysates. Means are from at least n=3, n=2, n=4 and n=2 independent experiments on cells from different patients, for pERK, pAKT, pFAK^{Y397} and cyclin D1, respectively. (E-H) Effect of Ritonavir at 1μM for 72 hours on the proliferation of schwannoma cells (NF2-/-) and Schwann cells (NF2+/+). (E-F) shows immunofluorescence analysis of a representative experiment. (E-F) Both were seeded in Lab-Tek chambers and cultured to a convenient confluence (>80%). Then they were incubated for 72 hours with ritonavir at a concentration of 1μM. After incubation, cells were fixed in 4% PFA, blocked, and incubated with anti-Ki67 antibody. A fluorescent-conjugated secondary anti-mouse antibody was used to reveal staining. DAPI was used to mark cell nuclei. (G) and (H) represent the percentage of Ki67+ cells per total counted cells (DAPI), normalized to the condition without any treatment. Means of Ki67+ cells for each condition (Untreated control, ritonavir 0.1 and 1μM) were obtained from n=4 and n=3 independent experiments, for schwannoma and Schwann cells, respectively.

8.1.6 Discussion

We found that HERV-K (HML2) is overexpressed in our tumours of interest, the schwannomas. However, we found a few schwannomas being negative for HERV-K (HML2) Env expression. Having a negative staining possibly reflects inactivity of HERV-K (HML2) in part of the population. However, Gag seem to be detectable in all patients (no negative staining). Without patients' genomic sequences, which HERV-K (HML2) sequence is responsible for protein synthesis in each patient cannot be predicted. In this context it cannot represent a universal tumour antigen. Even if HERV-K (HML2) seems present in many different tumours, usually at higher levels than

healthy controls, new tumours need to be checked for HERV-K (HML2) ORFs expression before thinking of using it as a general tumour-associated antigen.

Commercial HERV-K (HML2)-specific antibodies decreased proliferation as measured by ki67 staining, it is in accordance with a previous *in vivo* and *in vitro* studies on breast cancer by Wang-Johanning et al. (2012) who suggested apoptosis to be induced by their HERV-K (HML2) Env-specific antibody. Using commercial antibodies, we observed a decrease in some proliferative pathways as measured by immunoblots of p-ERK, p-AKT and Cyclin D1. Caspase 3 cleavage needs to be assayed in schwannomas treated with HERV-K (HML2)-specific antibodies, as an increase in the cleavage of caspase 3 would denote an engagement in apoptosis processes (Porter and Jänicke, 1999). Interestingly, HERV-K (HML2)-specific antibodies appeared to affect only schwannoma and not Schwann cells, with schwannomas harboring more HERV-K (HML2) proteins than their healthy counterparts as measured by western blots, immunohistological and immunofluorescent staining. This finding suggests the effect of HERV-K (HML2)-specific antibodies to be specific to the tumour, and so potentially a therapy targeting the tumour only.

Antiretroviral therapy has been considered for the treatment of some tumours in the past (Batchu et al., 2014; Gaedicke et al., 2002; Kumar et al., 2009). Here, we provide more supporting evidence to that idea. In addition to that, the lowest concentration tested *in vitro* that affected cell proliferation (0.1μM) as measured by Ki67 staining 72 hours post-treatment, is 2 times lower than the trough concentration in plasma from individual assessed 12 hours after drug administration (mean~0.21μM; range 0.17–0.26μM); the regimen consisted of the equivalent of 100mg ritonavir administrated twice daily for 7 days in pills formats that also contain lopinavir, another retroviral protease inhibitor (at a quantity equivalent to 400mg). Adverse effects were reported as being lower than grade 3 and 4, involving diarrhea, fatigue, nausea and headache; still the drugs were considered as well tolerated (Jackson et al., 2011). Also, in a previous study on HIV-infected patients conducted to improve drug uptake associated with less adverse effects, the group of patients that did not experience adverse effects, exhibited a trough concentration of 10.4μM (5.7–11.9μM), in the plasma, assayed 6 hours after a full regimen involving the administration of the following dosing scheme: 300mg

twice a day for 3 days, 400mg twice a day for 4 days, 500mg twice a day for 5 days and 600mg twice a day in combination with two nucleoside analogues, until the blood test (Gatti et al., 1999). In this case, the lowest concentration tested *in vitro* that affected cell proliferation (0.1μM) is 104 times lower than the trough concentration in the plasma. Deliver ritonavir in such a way could prevent or relevantly reduce adverse effects, while improving the life quality for patients contracting schwannomas.

In addition, HERV-K (HML2) protease was shown to resist ritonavir (very high inhibition constant (Ki) in comparison to HIV protease) (Kuhelj et al., 2001). The effect of the drug is unlikely related to HERV-K (HML2) inhibition, as the effect on proliferation measured through ki67 staining (~60% decrease) were really high in comparison to HERV-K protease activity measured by unprocessed gag polyprotein blotting (~30% increase of uncleaved Gag). In addition, ritonavir is known to affect the AKT pathway (Batchu et al., 2014) and the proteasome (Gaedicke et al., 2002), inhibitors to the latter are known to be anti-proliferative and induce apoptosis in tumor cells and have been used for cancer therapy (Crawford et al., 2011).

8.2 HERV-K (HML2) contribution to tumour development

8.2.1 Background

We found that HERV-K (HML2) is overexpressed in our tumours of interest, the schwannomas. However, we found a few schwannomas being negative for HERV-K (HML2) Env expression. Having a negative staining possibly reflects inactivity of HERV-K (HML2) in part of the population. However, Gag seem to be detectable in all patients (no negative staining). Without patients' genomic sequences, which HERV-K (HML2) sequence is responsible for protein synthesis in each patient cannot be predicted. In this context it cannot represent a universal tumour antigen. Even if HERV-K (HML2) seems present in many different tumours, usually at higher levels than healthy controls, new tumours need to be checked for HERV-K (HML2) ORFs expression before thinking of using it as a general tumour-associated antigen.

It is unknown whether HERV-K (HML2) is involved in schwannomas growth. A few lines of evidence suggest HERV-K (HML2) to contribute to tumour growth, as using RNA interference that knocks down HERV-K (HML2) *env* in breast cancer cell lines resulted in a decrease of cell proliferation and signaling protein activation such as p-ERK and Ras *in vitro*; the same observation was made *in vivo* in mouse engrafted with tumourigenic melanoma and breast cancer cell lines in which HERV-K (HML2) *env* was knocked down (Oricchio et al., 2007; F. Zhou et al., 2016). In addition, HERV-K (HML2) *env* overexpression resulted in activation of the signaling proteins p-ERK and Ras, but not in an increased proliferation *in vivo* in mice engrafted with breast cancer cell lines modified to overexpress HERV-K (HML2) *env* (F. Zhou et al., 2016). Also, HERV-K (HML2) *env* overexpression *in vitro* in HEK293T resulted in activation of p-ERK and several downstream transcription factors associated with cell transformation, although noticeable transforming activity could not be observed, in comparison to the JSRV *env* that induced both (Lemaître et al., 2017). Although, the full mechanism which could explain such effect remains unclear, the main signaling network involved have been suggested to be the Ras-MEK-ERK pathway, since that is affected by the HERV-K (HML2) knock-down and overexpression, and even if HERV-K *env* cannot cause tumour formation, it may contribute to the process.

Another example of a mechanism exists for *Betaretroviruses* (HERV-K (HML2) belongs to that genus of retroviruses) such as JSRV in the sheep. In fact, overexpressed JSRV Env is able to transform cells *in vitro* (Alberti et al., 2002; Liu and Miller, 2007, 2005). JSRV Env was shown to interact with the surface receptor HYAL2 and block the interaction between Hyaluronidase-2 (HYAL2) and the 'Recepteur d'Origine Nantais' (RON, also known as Macrophage-stimulating protein receptor – MST1R). Unblocked RON can freely stimulate the AKT pathway, likely driving cell transformation (Danilkovitch-Miagkova et al., 2003; Liu and Miller, 2007). Another mechanism, more speculative would involve a capacity observed for *Gammaretroviruses*. HERV-W Env possesses fusogenic ability and is transported with exosome, significantly increasing their uptake (Tolosa et al., 2012; Vargas et al., 2014). Exosomes are showed to be involved in tumours as they usually transport growth factors that are massively released, enabling cell interactions in the tumour (Azmi et al., 2014).

We tried to knock down HERV-K (HML2) for assessing whether HERV-K (HML2) plays a role in schwannoma biology, and design sequences to test whether it is feasible to specifically knock down each ORF for defining the precise involvement for each protein. Although we did not identify a functional shRNA sequence, we tested whether HERV-K (HML2) proteins could be released from schwannomas and potentially sustain main pathways involved in schwannoma biology.

8.2.2 Knock-down using manually designed sequences

Four shRNAs targeting HERV-K (HML2) LTRs, *gag* and *env* were designed manually, following some key rules. Two sequences, a and b, were design for position nt 949-973 and position nt 1020-1048 in the LTRs. The positions were chosen downstream of the transcription starting site at position nt 793 (Fuchs et al., 2011). One sequence, c, was chosen as spanning the transcription splicing sites which give rise to the HERV-K (HML2) *env* sub-genomic RNA, from genomic unspliced *gag-pro-pol* transcript. The last sequence, d, was from Origene, purchased from a set of four shRNAs designed against ERVK6, which is the locus HERV-K108. All the sequences were chosen as matching 100% of most HERV-K (HML2) sequences on a prepared alignment of 20 HERV-K (HML2) loci. Usually about 16-20 out of 20 were matched at 100%. The

alignments included the following loci: HERV-K113 (19p12b), HERV-K116 (1p13.1), HERV-K102 (1q22), HERV-K106 (3q13.2), HERV-K117 (3q27.2), HERV-K107 (5q33.3), HERV-K109 (6q14.1), HERV-K108 (7p22.1), HERV-K118 (11q22.1),HERV-K119 (12q14.1), HERV-K 12q13, HERV-K103 (10p22.1), HERV-K101 (22q11.21), ERVK4 (3q21.2), HERV-K115 (8p23.1a), HERV-K60 (21q21.1), HERV-K104 (5p13.3), ERVK19 (19q11), HERV-K 2q21.1. The free energy –ΔG for annealing, were calculated and the sequences a, b, and c were chosen so that they have a low ΔG (high in negative) when possible, along with less secondary structure as predicted by a tool predicting RNA secondary structures (http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html).

MCF-7 was chosen to test the 4 shRNAs as they express HERV-K (HML2) (F. Zhou et al., 2016). Transfected MCF-7 where checked under fluorescent microscope 2 days post transfection to assess whether transfection was successful (Figure 16B). Non-transfected MCF-7 cells were used as negative control. The cells were lysed 3 days post-transfection and HERV-K (HML2) knock-down was assayed by western blot using HERV-K (HML2) Env and capsid-specific antibodies. Out of the four sequences, the sequence targeting Gag-pol transcript, d, was the only one showing a slight decrease for both Env and capsid, in comparison to the non-transfected condition (Figure 16C and D). However, that effect was not reproducible as knock-down was seen once out of two replicates. So, none of the sequences a, b, c or d could consistently knock-down HERV-K (HML2) proteins after being introduced in cells for 3 days in culture.

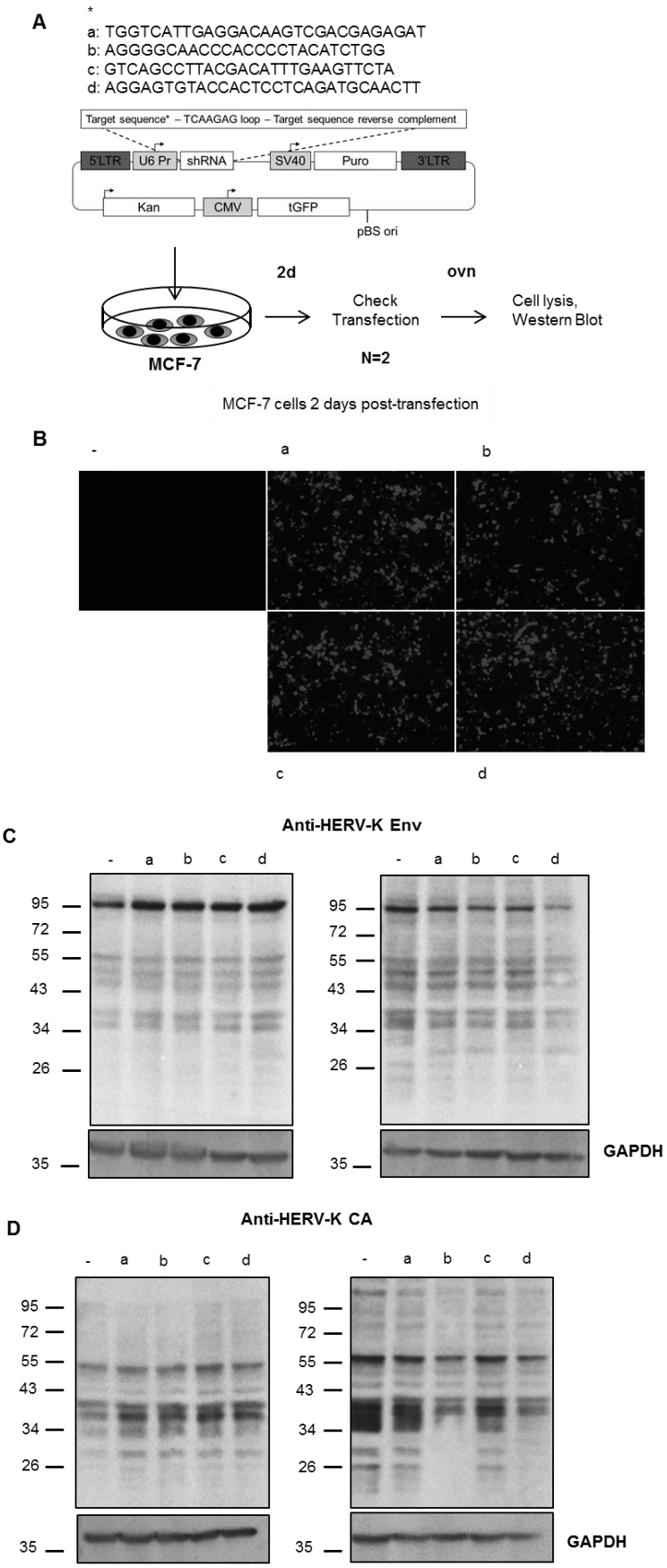


Figure 16 : HERV-K (HML2) knock-down experiments using Origene custom shRNAs.

Three home-designed and one company-designed shRNAs sequences were inserted in an expressing vector, each, possessing a GFP ORF as marker. MCF-7 cells were transfected *in vitro* using Fugene transfection reagent, followed by 3 days of culture and cell lysis for western blot. (A) a schematic representation of the experiment. Each shRNA sequence and the overall sequence of the vector are also indicated. At day 2, transfection efficiency was checked using fluorescent microscope (B). Cells were efficiently transfected. (C) and (D) represent duplicate runs of western blot, with membrane incubated with the commercial anti-HERV-K (HML2) Env (C) and anti-HERV-K (HML2) CA (D), respectively. GAPDH was used as a loading control for western blot. No reproducible decrease of HERV-K (HML2) bands was depicted. Blots are from duplicates from n=1 independent experiment.

8.2.3 Knock-down using a set of 4 sequences designed by the company – Origene

Home-designed sequences did not affect HERV-K (HML2) protein expression. Company-made lentiviruses expressing HERV-K (HML2) shRNAs (e, f, g, d) was used to perform a stable transduction of 293T cells. The lentiviral vector harbors puromycin resistance and GFP genes (Figure 17A). 293T cells transduced with lentivirus delivering scramble shRNA sequence was used as a control. After transduction, the cells were checked for GFP using a fluorescent microscope, and a day later, they were selected using puromycin at 4µg/ml. 5 days post-selection, a fraction of cells was lysed for western blot and the other fraction was counted and seeded at the same density into a bigger culture dish. 11 days post-selection, they were further counted. The same experiment was done on MCF-7, but transduction failed as no GFP was recorded and they did not survive puromycin selection at 4µg/ml (Figure 17B). Non-transduced MCF-7 lysate was included as a positive control as expressing HERV-K (HML2) Env. It is to be noted that MCF-7 lysate did not exhibit a band at 75kDa for the full-length Gag, which was present in 293T cell lysate (Figure 17C). Env and Gag blots showed a decrease only for the shRNA e in comparison to the scramble control, for the bands matching with MCF-7 lysate (Figure 17C). However, GAPDH blots used as a loading control revealed an uneven loading. Normalized to GAPDH, the decrease was not significantly different in Gag or Env for all shRNAs compared to the scramble control (data not shown). Interestingly, the 293T cells transduced with shRNA e seemed to proliferate less than the others as measured by the cell ratio between cell counts at day 11 and day 0 (Figure 17D). Also, cells transduced with shRNA f exhibit a higher cell proliferation (Figure 17D), along with a slight increase in HERV-K (HML2) Gag and Env expression, in comparison to the scramble control (Figure 17C). We noted that the

shRNA d had not affected HERV-K (HML2) expression in stably transduced 293T cells, questioning the observations made regarding transiently transfected MCF-7 (section 8.2.1). HERV-K (HML2) knock-down was not affordable using those sequences. Results are from n=1 independent experiment.

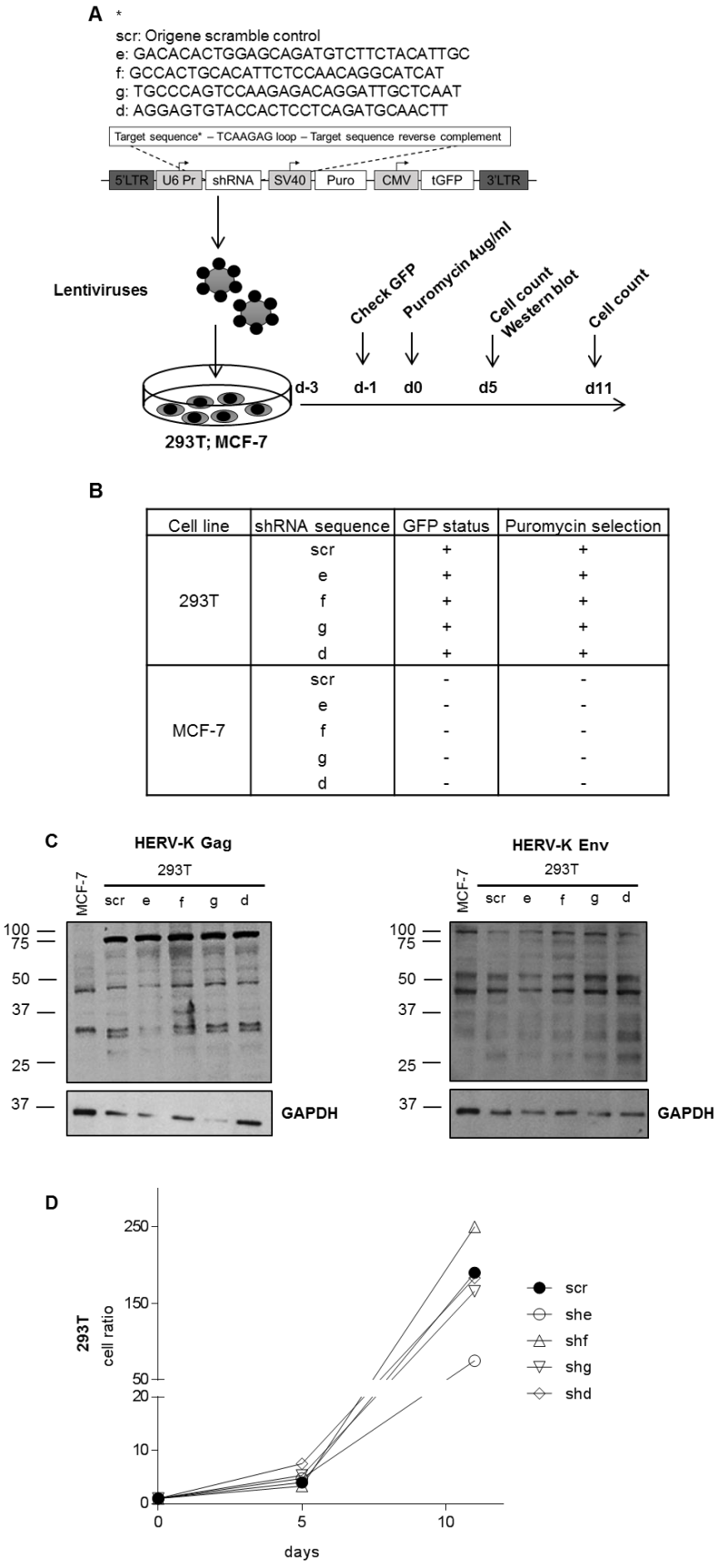


Figure 17 : HERV-K (HML2) knock-down experiments using Origene ready-made shRNA-expressing lentiviruses.

Company-made lentiviral particles expressing Origene-designed shRNAs sequences were used to transduce 293T and MCF-7 cells. The lentiviral vector possessed a GFP ORF as marker for assessing successful transduction. 293T and MCF-7 cells were incubated with the equivalence of 400 000 TU of lentiviruses in 2ml of culture medium, in presence of 8μg/ml of polybrene. 2 days post-transduction, GFP status was checked for each cell line. 3 days post-transduction, cells were selected using puromycin at a concentration of 4μg/ml. They were further cultured in selective-medium for 5 days, at that time they were counted and a fraction was used for western blot analysis; then further cultured for 11 days, when they were counted again. (A) Schematic representation of the experiment. Each shRNA sequence and the overall sequence of the lentiviral vector are indicated. (B) table summering the GFP status (transduced +, not-transduced -) and success of puromycin selection (survived selection +, did not survive -). (C) Represents western blot analysis using the commercial anti-HERV-K (HML2) Gag and anti-HERV-K (HML2) Env, respectively. GAPDH was used as a loading control for western blot. shRNA sequence e seems to decrease the amount of HERV-K (HML2) Gag but GAPDH is not even; and seems to decrease HERV-K (HML2) Env. (D) Graph that represents cell growth through cell ratios of cell number for day 5 and 11 of puromycin selection, in comparison to the number of cells seeded at day 0. shRNA sequence e showed the lowest cell ratio at day 11. Results are from n=1 independent experiment.

8.2.4 Knock-down using a set of 20 sequences designed using RNAi tool – Ui-Tei

Successful knock-down experiments from the literature described a siRNA and shRNA sequence targeting *gag* and *env*, respectively (Oricchio et al., 2007; F. Zhou et al., 2016). Interestingly, the design targeting *gag* knocked down *pol* and *env* transcripts, and ultimately decreased Env protein levels. As stated above, HERV-K (HML2) Env is produced from the splicing of *gag-pro-pol* genomic transcript. So, there is a caveat on understanding how *env* sub-genomic transcripts are triggered by shRNA targeting *gag* region. Because of this concern, I have designed 20 sequences targeting specific sites of HERV-K (HML2) genomic sequence using Ui-Tei RNAi tool: 5 sequences were selected in LTR downstream the TSS at position nt 793, 5 sequences were selected in the LTR upstream the transcription termination site at position nt 876 (Fuchs et al., 2011), by this mean all spliced forms are targeted; 5 sequences were selected in *gag* ORF and 5 sequences were selected in *env* ORF, this time not spanning the splicing site. Each LTR, *gag* or *env* sequence was submitted individually as query on the bioinformatics tools. 5 sequences were chosen randomly from the proposed sequences. Only the Tm was taken into consideration, as it is suggested that lower Tm, below 21°C, is associated with less off-target (Ui-Tei et al., 2008). All the sequences were introduced into 293T and MCF-7 cell lines through transient transfection. The cells were cultured

for 5 days after transfection, and the levels of effectively transfected cells were monitored via GFP signal using flow cytometry, alongside of that, a fraction of cells were lysed and knock-down was assayed using western blot (Figure 18A). For 293T, the transfection range was 33–47% GFP+ cells as compared to 13–22% GFP+ cells for MCF-7 (Figure 18B and C). Non-transfected 293T and MCF-7 cells were used as negative control. The samples corresponding to LTR targets (1-10) and *env* targets (16-20) were blotted with HERV-K (HML2) Env-specific antibody. The ones corresponding to *gag* targets were blotted with HERV-K (HML2) Gag-specific antibody.

Regarding 293T, among the sequences 1-10, the sequence 2 seemed to have decreased the amount of Env blotted on the membrane, as every bands seemed to decrease (Figure 18C). However, analysis of bands intensity normalised to the band of the loading control (GAPDH) revealed a non-significant change (data not shown). Among the sequences 16-20, the sequence 16 seemed to have slightly decreased the band at 55 kDa, but not the one at 38kDa. Analysis of bands intensity normalised to the loading control was in concordance with the observation (data not shown). Among the sequences 11-15, the sequences 12 and 13 seemed to have slightly decreased the bands from Gag at 45 and 50kDa, compared to the control (Figure 18B). However, this was not reliable as the band below (27-36 kDa) were denser than the ones in the control.

Western blots from transfected-MCF-7 were not reliable (due to technical issues, not shown). Altogether, the observed patterns suggested that no sequence was functionally interfering with the expression of HERV-K (HML2) Gag, or Env, or both. Also, the little number of transfected cells (less than 50%) makes it harder to estimate a decrease in HERV-K (HML2) proteins as less than 50% decrease in bands intensity would be expected if the shRNA was able to remove 100% of HERV-K (HML2) proteins in the transfected-cells. Results are from n=1 independent experiment.

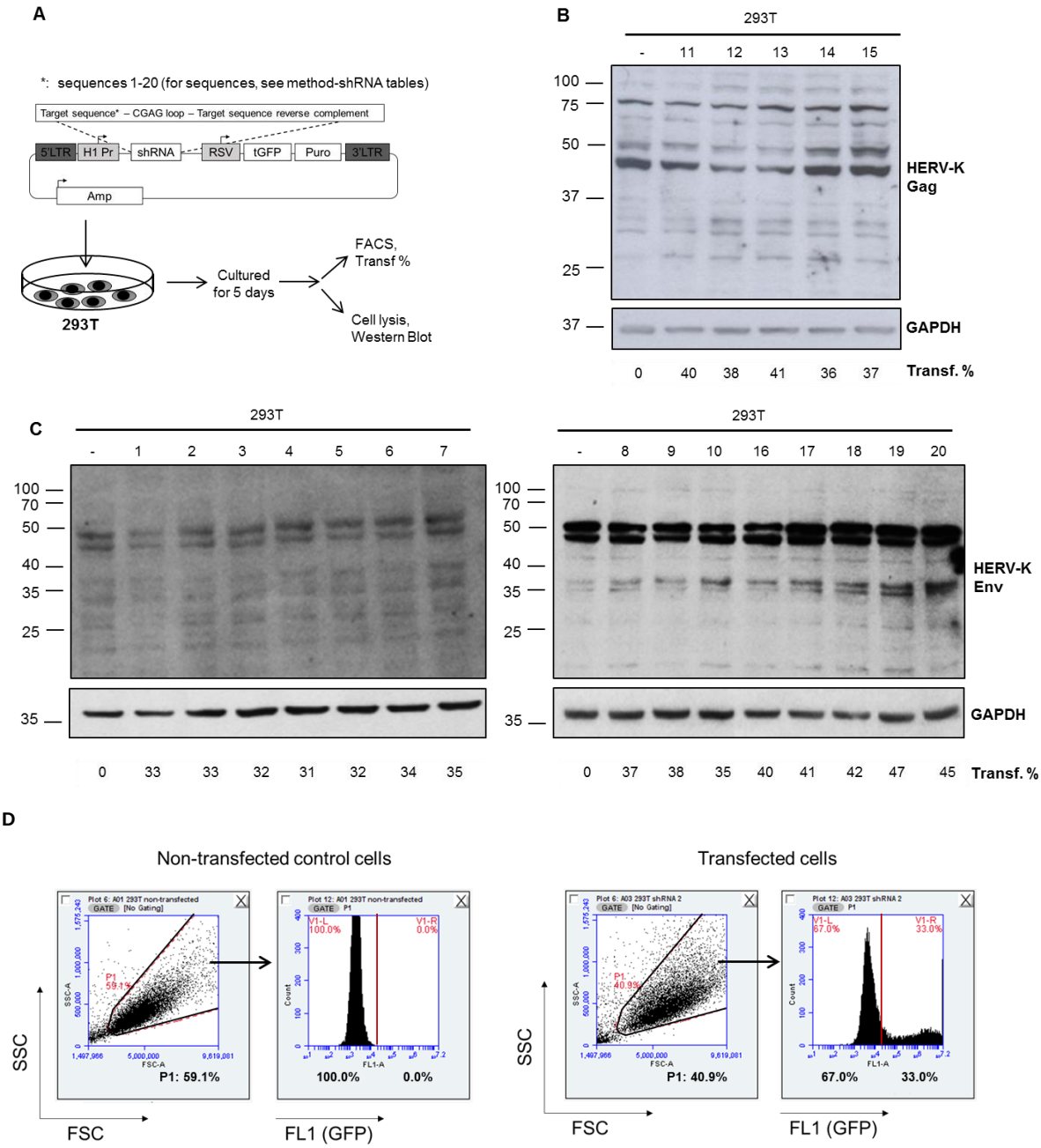


Figure 18 : HERV-K knock-out experiments using custom shRNAs from AMSBio.

LTR-specific (1-10; potentially knocking down both Gag and Env), Gag-specific (11-15) and Env-specific (16-20) shRNA sequences were introduced in lentiviral vectors by AMSBio. Each vector was used to transfect 293T cells, cultured for 5 days after transfection. A fraction of cells was used to monitor transfection efficiency by GFP expression using FACS, and another fraction of cells was used for western blot analysis. (A) Schematic representation of the experiment. (B) Shows western blot analysis of HERV-K (HML2) Gag expression in cells transfected with vector expressing shRNA 11-15, and not transfected as control. % of transfected cells (% of GFP+ cells) is indicated (Transf.%). (C) Shows western blot analysis of HERV-K (HML2) Env expression in cells transfected with vector expressing shRNA 1-7, and shRNA 8-10 and 16-20, and not transfected as control. % of transfected cells (% of GFP+ cells) is

indicated (Transf.%). (D) FACS gating strategy. Results are from n=2 independent experiment.

8.2.5 HERV-K (HML2) are released from schwannomas

HERV-K (HML2) proteins are overexpressed in schwannomas. Whether they play a role in the schwannomas remains unknown as designed shRNAs were not functional to address that fundamental question. Data from section 8.1 suggested that antibodies made against HERV-K (HML2) capsid and Env affect signalling pathways that are associated with proliferation and survival of schwannomas in culture. Since the main function of antibodies is to bind to their targets, it appeared reasonable to look whether HERV-K (HML2) proteins are present in the culture environment. Gag does not harbor any domain that could suggest that it locates at the membrane, so it should be accessible only if secreted. Env, by definition, should be found on the surface of the cells (Wang-Johanning et al., 2012). The hypothesis was that HERV-K (HML2) Gag and Env can be secreted. To test whether Gag and Env are freely released, cell-free medium from cultures of schwannoma (n=3) and Schwann cells (n=3) was harvested routinely and stored at -20°C. Medium were loaded on a gel for western blot purpose. HERV-K (HML2) Env and Gag-specific antibodies were used to detect the presence of Env or Gag in the medium. Culture medium that has not been in contact with any cells was used as a negative control, and cultures from MCF7 was used as this cell line is known in the literature to express HERV-K (HML2) (positive control). Gold staining was used as loading control, to assess the presence of proteins on the membrane. Env blots revealed a band arond 150 kDa, present in the negative control, and densely higher in schwannomas and Schwann cells (Figure 19A). Medium meant for cultures contained FBS, insulin, forskolin, β-heregulin, needed for the culture of schwannoma and Schwann cells. The antibody likely cross-react with components from medium preparation. The presence of the band at a very low density in the negative control could suggest that the HERV-K (HML2) Env-specific antibody cross-react with a ~150 kDa protein that is over produced in cultures. Part of that product, produced in culture, is possibly a full-length Env completely glycosylated (110-120 kDa), contributing to the band being denser. Also, Chen et al. reported detecting a 160 kDa protein in lysates from leukemia cell lines using a home-made HERV-K (HML2) Env-specific antibody, and suggested it to be the product from *pol-env* translation (Chen et al., 2013). Interestingly, that band was faint in medium from MCF7 culture. More specifically, a

band of ~38 kDa could be spotted in lanes of culture media of Schwann cells, not present in the negative control (Figure 19A). The band corresponds to the size of the glycosylated TM. That band were not found or less visible in the lanes of culture media from schwannomas and MCF7 cells.

Gag blots revealed a band at 150 kDa too, that is present at low density in the negative control, and in the media from Schwann cell cultures, but denser in media from schwannoma and MCF7 cultures (Figure 19B). The presence of the band at a high density in schwannomas' media suggest that the HERV-K (HML2) Gag-specific antibody recognised a ~150 kDa protein that may be further produced in culture, as previously with Env. The nature of that band would correspond to the Gag-pro-pol precursor protein (~160 kDa, see Figure 6, section 5.5). More interestingly, Gag blots also revealed band around 45 kDa, present in MCF-7 medium, likely corresponding to the p15+CA product, and only present in media from schwannoma cultures. This suggested that Gag can be found in medium from Schwannomas and not from Schwann cell cultures. Altogether, despite having a weak cross-reaction of the antibodies with protein from the medium preparation, these findings support the idea that HERV-K (HML2) proteins can be released from schwannomas and Schwann cells.

8.2.6 HERV-K (HML2) proteins are released in exosomes from schwannomas

Previous studies reported the secretion of HERV-W Env through exosomes (Tolosa et al., 2012; Vargas et al., 2014), guiding our work into checking whether HERV-K (HML2) proteins could be transported by exosomes. To this purpose, schwannoma primary cells were cultured into medium which had FBS replaced by commercially available exosome-free FBS. Medium from cultures was stored at -20°C and further processed when needed. Exosome processing consisted of mixing medium thawed from -20°C to exosome extraction reagent (see section 7.22), incubated overnight at 4°C and centrifuged at 10 000 x g for an hour. Pellets were resuspended in lysing buffer, ready to be run in a gel for western blot (Figure 19C, D and E). Anti-CD63 was used on blots as exosome-positive marker in the pelleted fraction from the medium. CD63 was presents on blots from the exosome fraction from schwannomas (lane 2, Figure 19C). It suggested exosomes are released from schwannoma cells. Alongside, schwannoma cell lysate obtained from the same cultures was used as cellular fraction, and as control

for band comparison between exosome and cellular fractions. It revealed a pattern for CD63 which was different but located around the same size, likely suggesting different levels of translational modification between the cellular and the exosome version of the protein. Envelop blots (Figure 19D) showed two dense bands at 55 kDa and 70 kDa and two faint bands at 38 kDa and 40 kDa, likely corresponding to glycosylated TM (38-55 kDa) and SU+TM (70 kDa). In comparison to the cellular fraction, the 98 kDa-glycosylated SU+TM band is missing. Gag blots (Figure 19E) revealed the 75 kDa-full length Gag, 57 kDa-p15+CA+NC, 38 kDa-CA+NC (weakly present in exosome fraction) and 27 kDa-CA (stronger in exosome fraction) bands, in both fractions. A band present between 75 and 57 kDa in exosome fraction, is absent in the cellular fraction. It may be an unspecific cross-reaction of Gag-specific antibody. Altogether, our data support the presence of HERV-K (HML2) Gag and Env in exosomes released by schwannoma cells as some of the HERV-K (HML2)-specific bands could be detected.

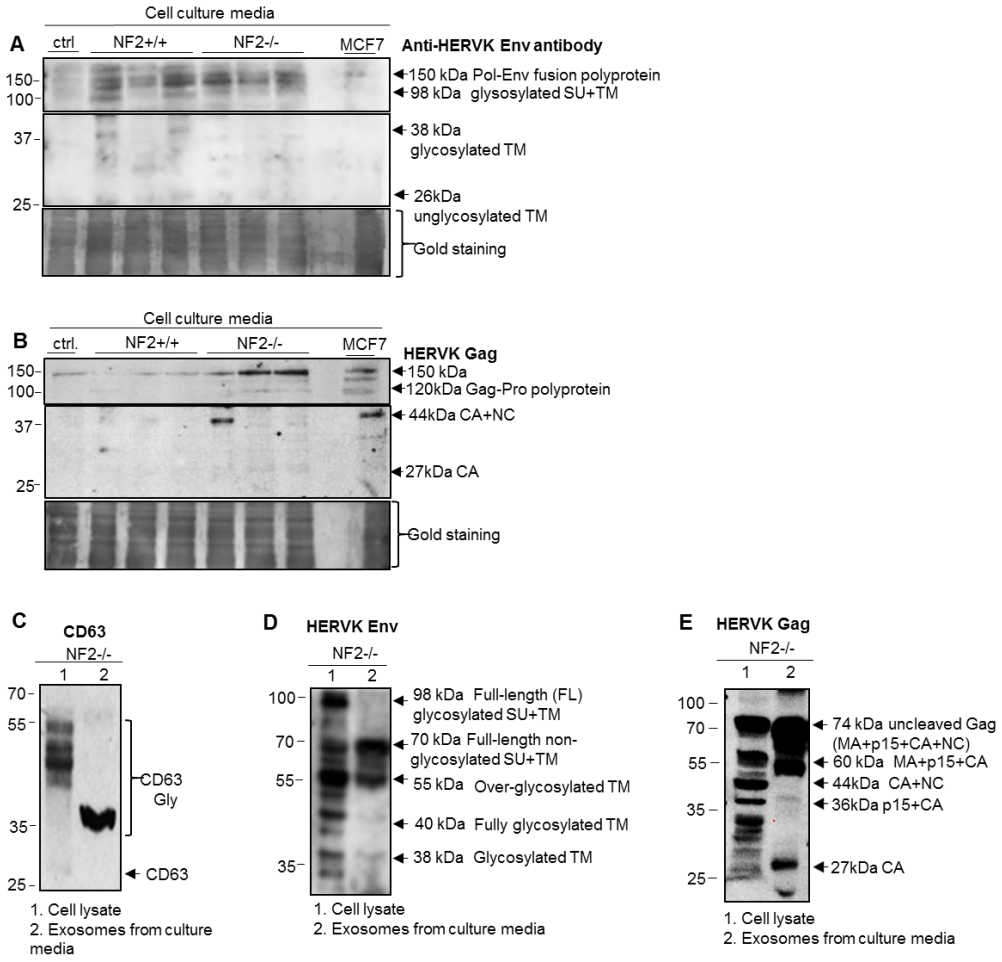


Figure 19 : Release of HERV-K (HML2) proteins from schwannomas and Schwann cells.

(A-B) Shows western blot analysis of HERV-K (HML2) Env and Gag expression in culture media collected from 3 separated primary Schwann and 3 separated primary schwannoma cell cultures. Gold staining is used as a loading control. Culture-free medium is used as negative control (Lane 'ctrl'). The presence of a weak band denoted a weak cross-reactivity of the anti-HERV-K (HML2) Env and Gag with a compound in medium composition at ~150 kDa. (C-D) represent western blot analysis of cellular (lane 1) and exosome (lane 2) fractions obtained from schwannomas from one out of n=3 independent experiments. (C) CD63 was blotted. (D) HERV-K (HML2) Env is blotted. Exosomes transport unglycosylated full length Env (70 kDa) and glycosylated TM (55 kDa, 40 kDa, 38 kDa). (E) HERV-K (HML2) Gag is blotted. Exosomes transport full length Gag (74 kDa), MA-p15-CA (60 kDa), CA-NC (44 kDa), p15-CA (38 kDa), CA (27 kDa).

8.2.7 HERV-K (HML2)-derived 9-mer peptides induced proliferation along with p-ERK, p-FAK, p-AKT, Cyclin D1, c-Jun

Schwannomas over-produce HERV-K (HML2) proteins and the use of HERV-K (HML2) capsid and Env-specific antibodies interfered with cell growth. Following the idea that antibodies cannot freely access intracellular space, the way by which HERV-K (HML2) proteins could be specifically targeted is by binding to proteins accessible in the extracellular environment. Two scenarios could be visualised: 1 - The antibodies bind to surface proteins which impact cell proliferation (Wang-Johanning et al., 2012). 2 – The proteins are released in the medium of cultures and high levels of proteins in medium stimulate cell proliferation, so that the use of antibodies binding to them neutralise such effect. To test whether it is plausible that proteins released in the extracellular space can impact cell proliferation, two commercially available 9-mer peptides from *gag* and *env* ORFs, respectively, were spiked at a concentration of 5μg/ml in growth factor-free medium after 24 hours of starvation. The cells were further kept for 72 hours to 6 days, then they underwent Ki67 and c-Jun staining for immunofluorescence or were lysed for western blot to look for Cyclin D1, p-ERK, p-FAK and p-AKT levels. Cells kept in growth factor-free medium free of peptides was used as a control. Both Gag and Env 9-mer peptides induced a significant increase of Cyclin D1, p-ERK, p-FAK and p-AKT levels measured 72 hours after peptide introduction into culture, in comparison with the control (Figure 20A, B, C and D). In addition to that, the number of cells stained positive for Ki67 and c-Jun, 6 days after peptide introduction into culture, was significantly increased (Figure 20E, F and G). This observation suggests that peptides from HERV-K (HML2) present in culture could have an impact on cell proliferation. Nonetheless, scrambled peptides designed using the amino acid composition from both Gag and Env peptide was not available, and would be critical in checking whether this effect is HERV-K (HML2)-specific.

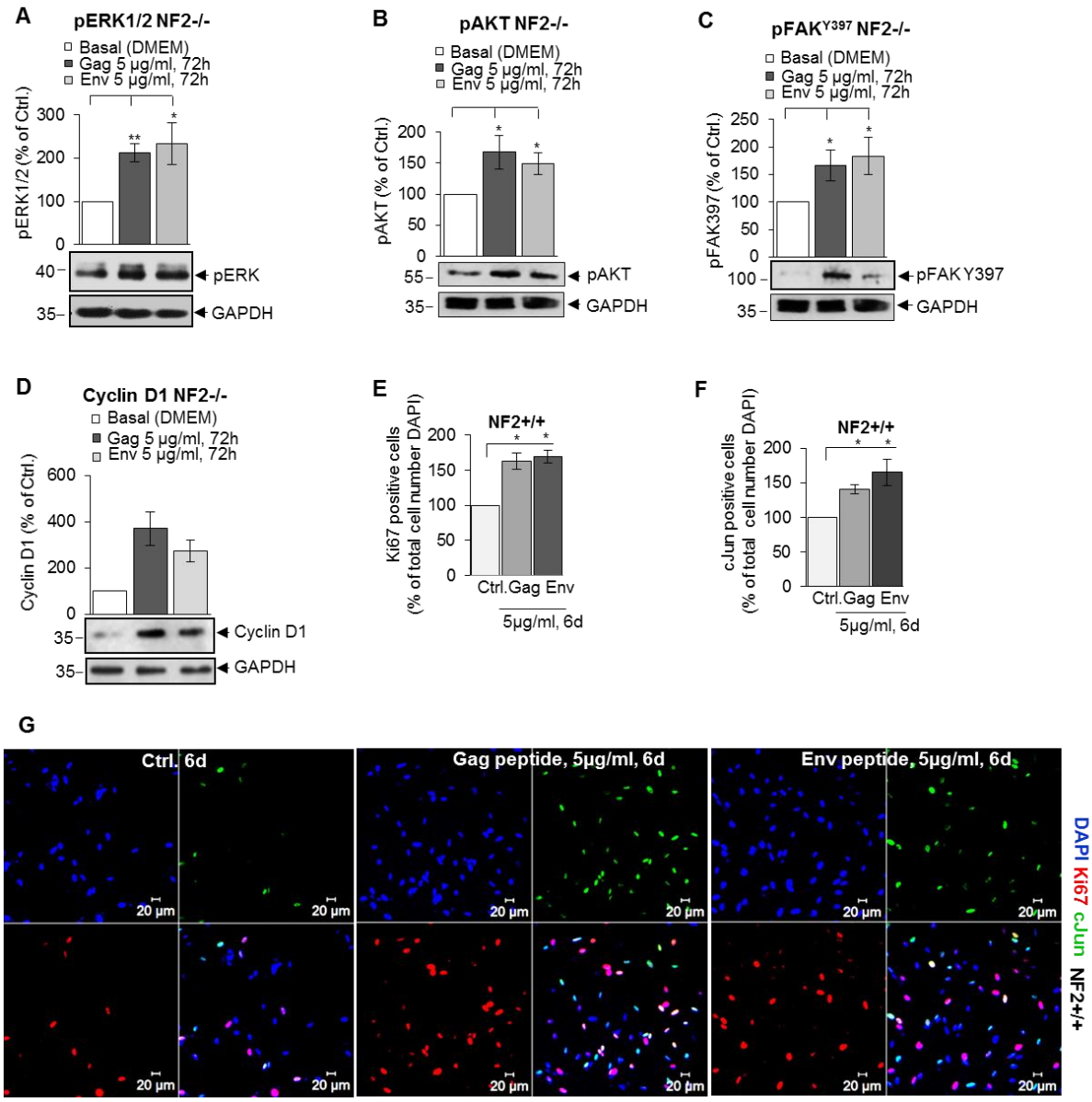


Figure 20 : HERV-K (HML2)-derived 9-mer peptides induced proliferation along with p-ERK, p-FAK^{Y397}, p-AKT, Cyclin D1, c-Jun.

Primary schwannomas (NF2-/-) or Schwann cells (NF2+/+) were starved for 24 hours. Then HERV-K (HML2) Gag-specific peptide, or Env-specific peptide, was introduced in culture at the concentration of 5µg/ml, and the cells were further kept for 72 hours and were lysed for western blot to look for Cyclin D1, p-ERK, p-FAK^{Y397} and p-AKT levels (A-D); or cells were kept with peptides for 6 days, then underwent Ki67 (red), c-Jun (green) and DAPI (blue) staining for immunofluorescence (G). Culture without adding any peptide was used as negative control. (A-D) show a significant increase of mean intensities of p-ERK, p-AKT, p-FAK^{Y397} and Cyclin D1, obtained from at least n=3 independent experiments. (E-F) represent the percentage of Ki67+ (E) and c-Jun+ (F) cells per total counted cells (DAPI), normalized to the condition without any peptide added. Means of Ki67+ and c-Jun+ cells for each condition were obtained from at least n=3 independent experiments using cells from three different patients and donors.

8.2.8 Discussion

I failed to identify a functional knock-down sequence for HERV-K (HML2), as none of the ones tested could confidently decrease HERV-K (HML2) proteins in cells. The sequences a, b and c were designed manually. None of them showed efficiency in HERV-K (HML2) knock-down. The sequences 1 to 20 were designed using RNAi tool which algorithm is based on the characteristics of functional siRNAs experimentally tested. None of them could knock HERV-K (HML2) down, suggesting that tool may not be powerful in RNAi design. Considering a range of available free RNAi tools to get sequence with the highest score in each of them could be a way forward.

One of sequences used in the past that successfully knock-down HERV-K (HML2) Env, had their target sequences in *gag-pro-pol* region (Oricchio et al., 2007), which is produced from a splicing of *gag-pro-pol* genomic transcript. These findings suggest interference could happen during splicing in the nucleus. In case the Env could be produced from genomic unspliced transcripts, knocking the genomic RNA down could reduce the expression of the Env. But it is unlikely. Our shRNA design was ambitious but would have been useful to confirm whether targeting *gag-pro-pol* knocks down the Env (Gag-pro-pol and Env protein knock-downs are somehow inter-connected), and decipher which proteins are involved or not in tumour growth, in case specific knock-down of each ORF is feasible.

HERV-K (HML2) Env appears to be released as HERV-K (HNL2) Env-specific antibody reacted with media from schwannoma and Schwann cell cultures. Interestingly, SK-N-SH neuron cell lines transfected with a DNA vector expressing HERV-K (HML2) *env* showed upregulation of Brain-Derived Neurotrophic Factor (BDNF) and Nerve Growth Factor (NGF) in the cells, and supernatant from those cultures, applied on another cell line (HFN cell line), induced BDNF and NGF in the latter cell line (Bhat et al., 2014). In addition to that, neural stem cells expressing HERV-K (HML2) *env* protects mice against neuro-inflammation. The mechanism is completely unknown, and other experiments showing similar results are highly needed. Also, whether HERV-K (HML2) Env is released and can stimulate cells, through a surface receptor, is to be tested. In our case, the band observed at a size of 150kDa seems too high to represent Env, because predicted size of fully N-glycosylated Env

would be approximately 120 kDa; while the band observed at a size of 38 kDa suggests that TM from Env can be released. The use of a 9-mer peptide derived from HERV-K (HML2) *env* in schwannomas culture after starvation induced pathways such as p-ERK, p-AKT, p-FAK and cyclin D1, and cell proliferation as measured by ki67 staining. This would suggest *env*-derived fragments has potentials in stimulating tumour pathways. Whether such effect is specific is to be tested by adding a scramble peptide with the same amino acid composition to the experiment. Also, HERV-K (HML2) *env* encodes a 699 amino acids protein, it would be very lucky that the 9-mer peptide used is in a functional domain of the protein. Interestingly the HERV-K (HML2) 9-mer peptides also induced c-Jun expression as measured by immunofluorescent staining. c-Jun is involved in demyelination, and is induced in Schwann cells in cases of injury, a condition that involves neuroinflammation (Parkinson et al., 2008). c-Jun is induced to avoid differentiation of Schwann cells, needed for proliferation which is required for nerve repair. In this manner, HERV-K (HML2) seems protective for neurons as suggest by (Bhat et al., 2014).

Regarding exosome-associated release of HERV-K (HML2) Env, it may be speculative to hypothesize that HERV-K (HML2) Env increase the uptake of exosome potentially produced by schwannoma cells. In fact, we do not have the tools to look for exosome uptake. However, such possibility is not to be ruled out, as it is the case for the Env of other HERV lineages, HERV-W and HERV-FRD that were shown to increase exosome uptake. In fact, knocking down HERV-W and HERV-FRD Env (syncytin-1 and syncytin-2) results in a decrease of exosome uptake (Vargas et al., 2014). Also, one HERV-K (HML2) locus, K108 (7p22.1), is estimated to be present in 100% of the population, and to date is the only one shown to possess a fusogenic capacity, as observed by ability of HIV particles pseudotyped with HERV-K108 *env* to enter into cells infected *in vitro* (Dewannieux et al., 2005). If HERV-K (HML2) facilitate exosome transactions, it would be a way of sustaining tumour development, since exosomes are involved in tumourigenesis (Azmi et al., 2014).

8.3 Possible mechanics for HERV-K (HML2) up-regulation

8.3.1 Background

During the time spent in the host genome, HERV-K (HML2) accumulated several mutations in its ORFs and LTR sequences, along with methylations, resulting in limited expression (Subramanian et al., 2011). Hypomethylation is the first requirement for HERV-K (HML2) transactivation (Hurst and Magiorkinis, 2017). In fact, the use of azacytidine, a drug that demethylates DNA, induces HERV-K (HML2) transcription (Depil et al., 2002). However, in some cell lines, demethylation is not sufficient to induce HERV-K (HML2) transcription (Fuchs et al., 2011; St Laurent et al., 2013), stressing that transcription factors need to be considered in HERV-K (HML2) transactivation. In fact, cell lines with similar genetic background, which differ by the expression of particular transcription factors would exhibit different levels of HERV-K (HML2) transcription (Fuchs et al., 2011). A list of transcription factors predicted *in silico* is available (Manghera and Douville, 2013). Only seven of them have experimental evidence for their involvement in HERV-K (HML2) transcription: OCT4 (Grow et al., 2015), NF-κB and NF-AT (Gonzalez-Hernandez et al., 2012), MITF-M (Katoh et al., 2011), Sp1 and Sp3 (Fuchs et al., 2011), YY1 (Knössl et al., 1999).

The hippo pathway regulates cell proliferation, death and differentiation. The core signaling cascade of the hippo pathway involves the phosphorylation of large tumour suppressor kinases 1/2 (LATS 1/2), which then phosphorylates Yes-associated protein (YAP) and transcriptional coactivator with PDZ-binding motif (TAZ); and phosphorylated YAP/TAZ interacts with the protein 14-3-3 that results in YAP/TAZ cytoplasmic retention. Moreover, YAP and TAZ phosphorylation leads to their degradation. When dephosphorylated, YAP/TAZ enters the nucleus and induce gene transcription by interacting with transcription factor TEAD (Yu and Guan, 2013).

CRL4-DCAF-1 is a protein ubiquitin ligase that get activated when Merlin is lost (L. Zhou et al., 2016), and which was recently identified as one of the regulators involved in the hippo pathway (W. Li et al., 2014). In the absence of Merlin, CRL4-DCAF1 inhibits LATS 1/2 that normally phosphorylates YAP/TAZ. When dephosphorylated, YAP/TAZ binds to TEAD and activate the transcription of oncogenic genes (W. Li et

al., 2014). In the present work, the objective is to test whether Merlin can influence HERV-K (HML2) transcription and whether CRL4-DCAF-1 and ultimately TEAD is involved in driving HERV-K (HML2) transcription when Merlin is lost. Also, for the first time, I propose TEAD as a potential transcription factor binding to HERV-K (HML2) LTR by *in silico* analysis of LTR sequences. The hypothesis was that Merlin loss induces the formation of YAP-TEAD transcription factor complex that can transactivate HERV-K (HML2) by binding to its LTR.

8.3.2 Merlin re-introduction influences HERV-K (HML2) proteins levels

Merlin loss is a key feature for schwannoma development, so we asked whether Merlin keeps HERV-K (HML2) expression under control. To gain some insights, Merlin was re-introduced into the tumour using adenovirus harboring the ORF that encodes for a functional Merlin, and HERV-K (HML2) protein levels were monitored in lysates from schwannoma cells collected after 3 to 5 days post-infection. Lysates from cells infected with an adenovirus harboring GFP only were used as control. An additional lane with lysates from non-infected cells was included in the experiment, in case of unwilled cross-reaction of HERV-K (HML2)-specific antibodies with adenoviral products massively introduced in the cell. In fact, a dense band sometimes appeared only in infected conditions, around 26 kDa (Figure 21B and D). For that reason, 27 kDa-TM for Env and 27 kDa CA for Gag were not assessed in the density analysis. In all experiments, Merlin was re-introduced successfully (Figure 21B and D, bottom panel). Merlin-Adenoviruses significantly reduced HERV-K (HML2) Gag and Env proteins in comparison to GFP-Adenoviruses, as measured by band densities (Figure 21C and E). All bands from 38-98 kDa for the Env and 38-75 kDa for Gag were significantly reduced in lysates from Merlin-Adenovirus-infected cells (Figure 21C and E). The densities plotted on the diagrams are means from n=6 and n=4 experiments with samples from different individuals, for Gag and Env, respectively. These observations strongly support the idea that Merlin keeps HERV-K (HML2) protein synthesis under control, at least partially, as HERV-K (HML2) was still detectable to level that represent more than half of the levels observed in lysates from GFP-adenovirus-infected cells (Figure 21C and E). However, whether it is direct or indirect remains to determine.

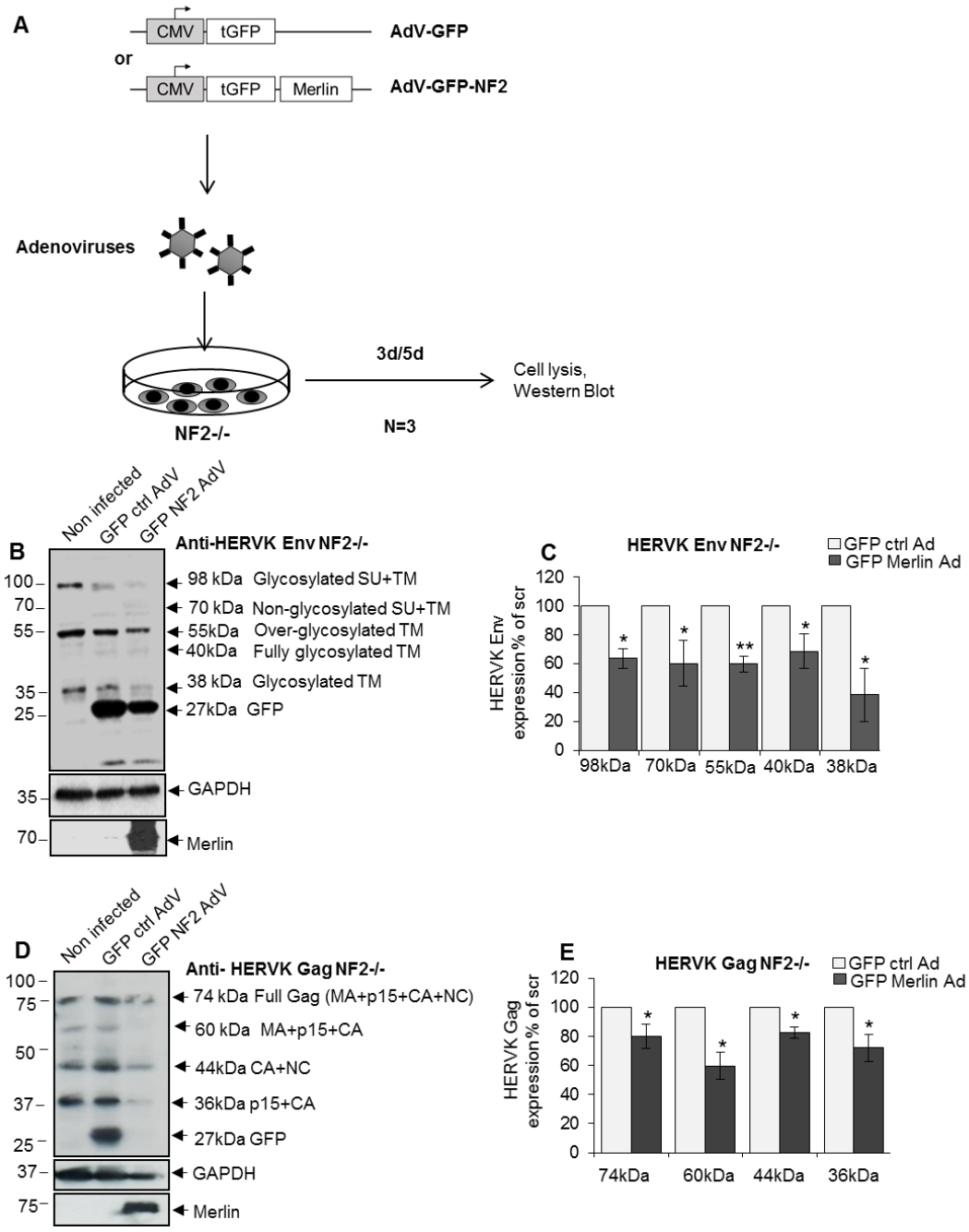


Figure 21 : Introduce a recombinant Merlin in schwannomas through AdV system results in a significant decrease of the amount of HERV-K (HML2) Gag and Env proteins.

AdV-GFP or AdV-GFP-NF2 were used to infect schwannoma primary cells. A non-infected condition was added to experiment. Cells were kept in culture for 3-5 days for analysis of Env and Gag by western blot. GAPDH was used as loading control for western blot. Merlin levels were assessed to ensure Merlin introduction. (A) Represents the experimental procedure. (B) and (D) show a representative western blot analysis for Env and Gag, respectively. A band around 27kDa appeared only in infected conditions, and so denoted a possible cross-reaction between anti-HERV-K (HML2) Env and Gag and components of AdV. (C) and (E) show mean intensities for each HERV-K (HML2) Env or Gag band normalized to GAPDH, normalized and represent as a percentage of

the mean intensities for each HERV-K (HML2) Env or Gag band in the condition infected with AdV-GFP. Mean intensities were obtained from n=4 and n=6 independent experiments using cells from different patients, for Env and Gag, respectively.

8.3.3 CRL4-DCAF-1 is unlikely involved in HERV-K (HML2) activation

Since CRL4-DCAF-1 was recently described as one regulator of the hippo pathway, downstream of Merlin (L. Zhou et al., 2016), we investigated whether CRL4-DCAF-1 take part in HERV-K (HML2) protein over-expression in the absence of Merlin. To this goal, schwannoma cells in cultures were transduced with lentiviruses expressing CRL4-DCAF-1-specific shRNA. Cells transduced with lentiviruses expressing scramble shRNA was used as control. CRL4-DCAF-1 was successfully knock down in schwannomas transduced with CRL4-DCAF-1-specific shRNA (Figure 22A). However, western blot analysis of HERV-K (HML2) Env was inconclusive as 98 kDa-glycosylated SU+TM band was decreased, 70 kDa-SU+TM was increased, 55 kDa-over-glycosylated TM was reduced, and 40 kDa and 38 kDa-glycosylated TM bands were not significantly affected (Figure 22B). The results are from n=3 independent experiments using cells from different patients. This observation suggested that CRL4-DCAF-1 is not involved in HERV-K (HML2) over-expression as products from Env, altogether, did not decreased significantly.

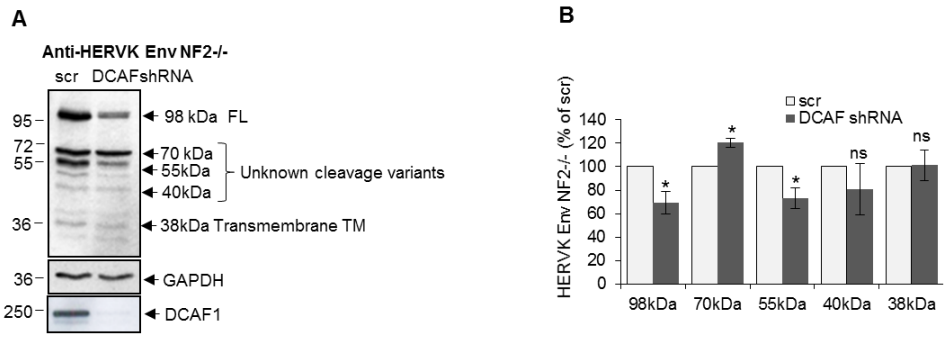


Figure 22 : Knock-down of CRL4-DCAF1 does not consistently affect HERV-K Env protein levels.

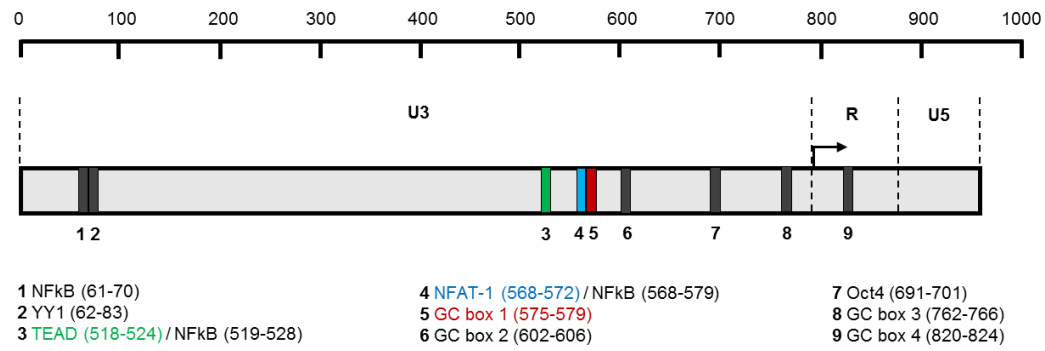
Primary schwannomas were transduced with lentiviral particles expressing DCAF1 shRNA. 72 hours post-transduction, the cells were selected in puromycin for 6 days, after what they were lysed to analyse HERV-K (HML2) Env expression through western blot. Infection with lentiviral particles expressing a scramble shRNA sequence was used as control. (A) shows a representative western blot analysis. GAPDH was used as a loading control. DCAF1 was blotted to assess successful knock-down. The intensity of each HERV-K (HML2) Env band was monitor in both infected conditions, lentiviruses expressing scramble and DCAF1 shRNA. (B) represents mean intensities for each band normalized to GAPDH, and presented as a percentage to the mean intensities obtained with the scramble control.

8.3.4 TEAD as a potential, novel transcription factors for HERV-K (HML2) LTR

HERV-K (HML2) LTR represents a land for several transcription factors. Binding motifs for 39 transcription factors have been predicted (Manghera and Douville, 2013). But, only seven have been tested experimentally: OCT4 (Grow et al., 2015), NF-κB and NF-AT (Gonzalez-Hernandez et al., 2012), MITF-M (Katoh et al., 2011), Sp1 and Sp3 (Fuchs et al., 2011), YY1 (Knössl et al., 1999). The figure 23 represents the binding sites predicted by Manghera et al. that were shown to bind to LTRs using ChIP and EMSA. Some NF-AT and NF-κB sites shown to be bound by their respective transcription factors (Gonzalez-Hernandez et al., 2012) are not depicted on the LTR map here as neither I, nor Manghera et al. found them. Oct4 has been introduced more recently as a result from extensive study of HERV activity in embryonic cells (Grow et al. 2015), and is depicted on the map. By mapping the binding sites proposed by Gonzalez-Hernandez et al. (2012) (using ALGGEN-PROMO and TRANSFAC software tools), Manghera and Douville (2013) (using ALGGEN-PROMO software) and I (checking manually the sequences on proviruses alignment based on putative binding motifs described in the literature), I hope to provide the reader with a synthetic and comprehensive mapping of transcription factors for HERV-K (HML2) LTRs.

In this work, a focus on conservation was also attempted. To this purpose, alignments of 19 HERV-K (HML2) provirus was used to find how conserved are the binding sites among LTR sequences (the number of LTR sequences that do not possess a single mutation in the binding motif). It came up that binding motifs in figure 23 were 100% conserved as following: (1) NF-κB in 17/19, (2) YY1 in 15/19, (3) NF-κB in 17/19, (4-5) NF-κB in 16/19, (5) GC box 1 in 16/19, (6) GC box 2 in 17/19, (8) GC box 3 in 18/19, (9) GC box 4 in 15/19, and (7) Oct4 in 12/19. In this context, I have been interested in TEAD as a potential transcription factor for HERV-K (HML2) LTR. TEAD binds to DNA consensus sequence TGGAAT (Vassilev et al., 2001). Looking manually for that canonical sequence on LTR, I found only one site, conserved at 100% (not a single nucleotide difference among LTR sequence queries) in 17/19 LTR sequences. Interestingly, TEAD binding motifs clashes with NF-κB on site (3) as they share 5 nucleotides in common out of 6 from TEAD (GGAAT). Based on this, testing whether TEAD could influence HERV-K (HML2) transcription was suggested. As

CRL4-DCAF-1 knock down did not relevantly affect the expression of HERV-K (HML2) Env, the use of YAP-TEAD inhibitors appeared to be a good way to further clear whether HERV-K (HML2) transcription in tumour could be driven by hippo-pathway via YAP-TEAD, independently of CLR4-DCAF1 (see section 8.3.4).



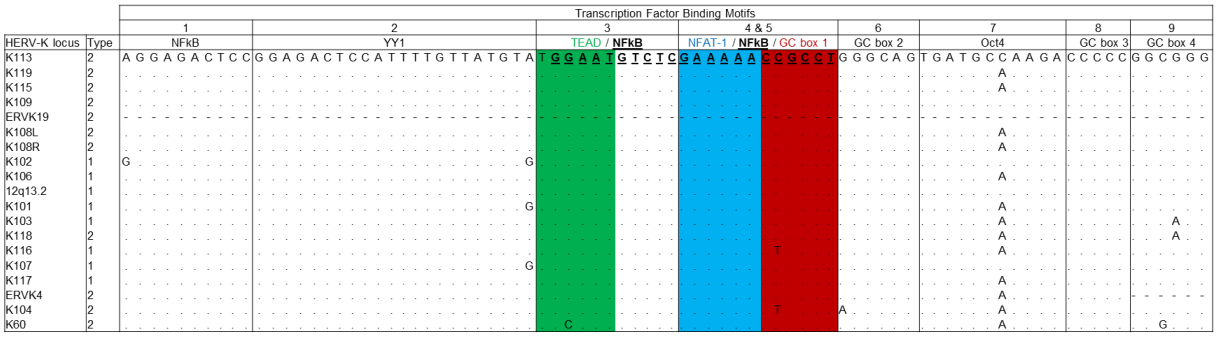


Figure 23 : The presence of TEAD binding motif on HERV-K (HML2) LTR *in silico*.

The transcription factor TEAD that binds to DNA consensus sequence TGGAAT (Vassilev et al., 2001), was manually searched and marked, it is fully conserved in 17/ 19 of the sequences aligned.

8.3.5 The effect of verteporfin, YAP-TEAD inhibitor, on HERV-K (HML2) transcription

MCF-7 cells were used to investigate the relation between YAP-TEAD pathway and HERV-K (HML2) expression. Verteporfin, a drug that disrupt the interaction between YAP and TEAD (Brodowska et al., 2014) was used at a final concentration of 4µg/ml. Both RNA and proteins were extracted after 48 hours incubation with both drugs. Primers against HERV-K (HML2) *pol* and *env* were used to measure HERV-K (HML2) transcription, relative to GAPDH. CTGF-specific antibody was used on western blot to assess the decrease of the amount of CTGF as a measure for drug efficacy (Figure 24, bottom panel). In fact, *CTGF* (connective tissue growth factor) is a target gene of the YAP-TEAD transcription factor complex. The expression of HERV-K (HML2) *pol* was not significantly affected, whereas *env* was reduced by third (three times less compare to DMSO control condition) following 48-hour treatment (Figure 24, top and middle panel). The data suggest that HERV-K (HML2) transcription can be regulated by YAP-TEAD transcriptional complex.

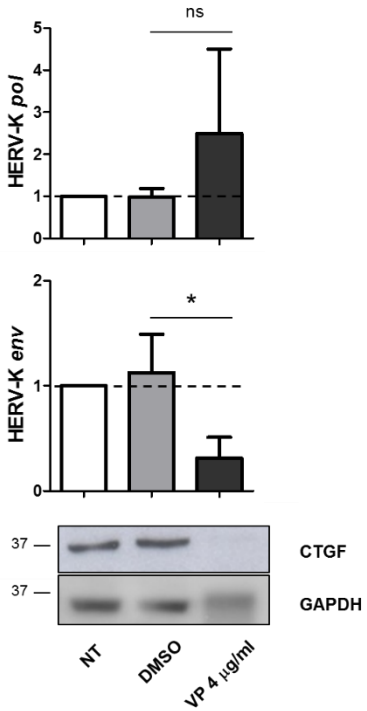


Figure 24 : YAP-TEAD inhibitor affects HERV-K (HML2) transcription.

MCF-7 cells were treated with verteporfin (VP, 4µg/ml), (DMSO as vehicle, negative controls), or left untreated for 48 hours, after what a fraction of cells were used for either western blot or RT-qPCR. Top panel represent *pol* and middle panel represent *env* expression as assayed by RT-qPCR. Bottom panel shows a representative blot of

CTGF (assessed for drug efficacy) and GAPDH (loading control). The data were gathered from n=3 independent experiments.

8.3.6 Discussion

A lot of pathways are dysregulated downstream of Merlin. One is CRL4-DCAF-1 along with the Hippo pathway. Our results suggested that CRL4-DCAF-1 is not the main partner by which Merlin could indirectly affect HERV-K (HML2). In fact, knocking down CRL4-DCAF-1 resulted in a non-significant decrease of HERV-K (HML2) Env, if any.

I also addressed whether inhibiting Hippo pathway using inhibitors for YAP-TEAD interaction, could impact HERV-K (HML2) expression. The use of verteporfin significantly affected HERV-K (HML2) *env* transcription. However, for all three experiments, the expression of GAPDH was also decreased by the treatment: in *env* RT-qPCR, Ct values were 26.4, 25.0, 25.8 for verteporfin condition compared to 22.2, 22.9, 22.4 for DMSO control condition. The same amount of total RNA (100ng) was used per condition. It suggests that verteporfin generally affects the intracellular levels transcription (mRNA). Similar observations were made for the loading control using western blot. *In vitro*, verteporfin was shown to induce a decrease of the expression of Cyclin D1, D3, A2, E1, Oct4, VEGFA, c-myc, Survivin, along with decreasing proliferation post-treatment (Brodowska et al., 2014). The knock-down of YAP and TEAD is required to further assess whether HERV-K (HML2) transcription is specifically driven by YAP-TEAD complex.

HERV-K (HML2) *pol* was increased (by half) post-treatment, not significantly. TEAD was described to be involved in gene repression when associated with co-repressor complex NuRD (Kim et al., 2015). Interestingly, NuRD is part of the machinery complex that could possibly limit HERV-K (HML2) transcription, as KAP1, a potent repressor of ERV expression, may recruit NuRD complex (Rowe and Trono, 2011). In this manner, TEAD could be a transcription factor for repressing HERV-K (HML2) transcription, and disrupting YAP-TEAD complex could indirectly release inhibition, promoting HERV-K (HML2) expression.

Finally, TEAD and NF-κB are sharing binding sites. NF-κB at that position has been shown experimentally to induce HERV-K (HML2) transcription. It is hard to predict whether there is a competition between both transcription factors to bind to the LTR. But, if it is the case, then NF-κB is likely the one which accesses the site as suggested by the literature. However, in schwannomas, the use of NF-κB inhibitor SN50 did not affect the expression of HERV-K (HML2) Gag (data not shown). NF-κB knock-down experiments in schwannomas are needed to test whether NF-κB is involved in HERV-K (HML2) transcription in schwannomas. Ultimately, Chromatin-immunoprecipitation need to be performed to find whether TEAD physically binds to the LTR of HERV-K (HML2).

8.4 The correlation between a recently integrated ERV in macaques and interferon stimulated genes (ISGs) during viral insult modeled by SIV infection

8.4.1 Background

Endogenous Retroviruses (ERVs) are descendants of ancient retroviral infections which have become established in the germline and proliferated so they now represent ~5% of the genome of humans and other mammals (more if we include the older group called MaLRs, Mammalian apparent LTR-Retrotransposons). Individual proviruses are termed loci and during their Mendelian transmission they have accumulated mutations over time which have rendered most of them replication-defective, with few loci retaining full-length Open Reading Frames (ORFs) for all genes and thereby the possibility of yielding infectious, cell-free progeny virions (Jern and Coffin, 2008; Stoye, 2012). ERVs, and associated retroelements such as LINES (Long Interspersed Nuclear Elements), are increasingly no longer regarded simply as 'junk DNA' given their abundance in host genomes (Kim et al., 2012). For example ERVs have been implicated in epigenetic gene regulation (Brattås et al., 2017; Fasching et al., 2015; Lavie et al., 2005; Rowe et al., 2010), and there are many examples of individual ERV loci being co-opted by the host, e.g. HERV-E LTR-driven tissue-specific expression of a human salivary amylase gene (Ting et al., 1992), ERV-derived syncytins that play a role in placentation (Lavialle et al., 2013) and the promoter-containing Long Terminal Repeats (LTRs) of ERVs that play a role in gene control (Lu et al., 2014).

Recently, ERVs and other retroelements have also been implicated in innate sensing and PAMP (Pathogen Associated Molecular Patterns) recognition (Chuong et al., 2016; Hung et al., 2015; Stetson et al., 2008). Indeed, Volkman and Stetson (2014) have suggested that retroelements themselves are a source of PAMPs, such as cytosolic DNA resulting from reverse transcription, that influence the activation threshold required for triggering the innate immune response. They speculate that the level at which activation occurs is an evolutionary trade-off between avoidance of tolerance of exogenous pathogens on one hand and the need to avoid constant triggering by endogenous elements on the other. The latter appears to occur in Aicardi-Goutieres Syndrome, where a defective enzyme that metabolizes endogenous nucleic acid leads to a build-

up of retroelement PAMPs such as cytosolic DNA that may trigger an inappropriate innate response (Stetson et al., 2008).

To evaluate the biological significance of ERV expression in an *in vivo* situation we have employed the macaque model, examining how an ERV and a key gene of the interferon signaling pathway behave under the pressure of an exogenous agent, in this case pathogenic variants of simian immunodeficiency virus (SIV). We hypothesized that individuals with elevated ERV transcription would have a higher interferon response in the presence of an exogenous, infectious agent. Differences in ERV expression levels between individual macaques are likely because recently integrated loci will exist in the host population in three states: full-length proviruses, solo LTRs (a relict, non-coding recombinant form) and 'empty' pre-integration sites, *e.g*. as shown in humans with the recently integrated HERV-K loci (Subramanian et al., 2011). Macaques represent a suitable model to study ERV expression *in vivo* in a number of key respects, e.g. their evolutionary and anatomical relationship to humans, having diverged about 25 million years (Richard A. Gibbs, 2007; Stocking and Kozak, 2008), has made them an established model to study human infectious disease and pathogens, including HIV/AIDS. The study of primates such as macaques also broadens out the wealth of data generated from mouse studies, which show innate signaling responses to ERV activity in some strains (Hurst and Magiorkinis, 2015). For example, there is no conclusive evidence that any human ERVs are replicating today (Bhardwaj et al., 2014; Jern and Coffin, 2008; Karamitros et al., 2016; Stoye, 2012), whilst some mouse ERVs are definitely replication-competent.

We selected PcEV (*Papio cynocephalus* Endogenous Retrovirus), one of only three ERV lineages that have been copying within the macaque genome in the last 5 million years (Magiorkinis et al., 2015), and STAT1, which is both an essential component in the type I interferon signaling pathway and is itself an Interferon Stimulated Gene (ISG) (Gough et al., 2012), and whose expression is elevated (Ferguson et al., 2014). In fact, one of the major targeted ISG is STAT1 (Der et al. 1998; Lanford et al. 2006) and its abundance, which is set by the basal level of IFNβ, defines the susceptibility towards infections (reviewed in Gough et al. (2012)). Therefore, monitoring the expression of STAT1 provides a good read-out of type I IFN response. We measured gene expression

levels in tissues that are key targets of infection by SIV during the acute period of infection, when SIV plasma levels are at their highest, (Ferguson et al., 2014), reasoning that this is where and when *de novo* innate signaling responses are likely to be stimulated.

We also examined the macaque reference genomes sequences to assess the transcriptional, translational and replicational capacity of its PcEV loci. ERVs do not need to produce full-length proteins to be detected by the immune system, e.g. antibodies to two defective HERV-K loci were found in 29% of healthy individuals (Hervé et al., 2002) and a CTL response to a pseudogene derived from another ERV lineage – HERVK (HML6) – was found in a melanoma patient (Schiavetti et al., 2002). However, the likelihood of an ERV locus being biologically relevant increases if its sequence is free of obvious inactivating mutations.

Here, we show the presence of multiple protein-coding but not replication-competent PcEV loci in the macaque reference genomes and describe the presence of PcEV RNA at low levels in the plasma in some infected macaques, which possibly represents circulating virus. Critically, we find that tissue-associated PcEV and STAT1 expression are strongly positively correlated, and this occurs almost irrespective of the level of SIV replication in the same tissue or in the plasma. The significance of this novel finding of a correlation between ERV activity and innate signaling are discussed.

8.4.2 Bioinformatic analysis suggests PcEV protein production but not replication.

An online search of the most recent rhesus macaque genome assembly (rheMac8) using our constructed PcEV reference sequence found 72 PcEV loci (plus another 76 matches in unassembled regions), most of which are represented only by fragments. In figure 25 we show the only 10 loci that do not have large regions (more than several hundred nucleotides) missing in the reference sequence. Among these loci we find examples of full-length ORFs for all genes but no single locus that has full-length ORFs in all genes. Examination the LTRs of the above loci (excluding the two older ones), showed them all to have unmutated TATA boxes (TBP binding sites) plus several transcription factor binding sites (table 4) suggesting transcription of these loci.

From these observations, we might therefore predict expression of all ERV proteins but not ERV replication. However, replication-competent loci might have been missed for several reasons. Firstly, there are many problems with the assembly of the macaque reference genomes. Four of the 10 most intact loci contain scaffold gaps, where sequenced regions either side have not been joined up. Locus chr9:50301533-10531 provides a good example of these problems. In the rheMac8 assembly, this locus appears to have its *pro-pol* reading frame interrupted by a single premature stop codon and a long insertion. This insertion consists of a tandem duplication at a region marked as a scaffolding break on the UCSC Genome Browser. Actually, neither the premature stop codon nor the insertion derive from this locus. We were able to determine this by examining the locus sequence from the earlier assembly of the rhesus genome, which comes from the same individual animal, and that of the homologous locus in the cynomolgus macaque genome, a species which diverged from the rhesus macaque ~1 million years ago (mya) (Jiang et al., 2016). By searching for the unique chimaeric sequences consisting of the end of the LTR and the contiguous host flanking region, we were able to find and extract the sequences of this locus in the rheMac2 build (chr9:56546894-55724) and in the macFas5 build of the cynomolgus macaque (chr9:55063875-78763). These three *pro-pol* sequences are not homologous along their entire length with a readily visible transition – a 'breakpoint' (near to position 3937 in our reference genome): the first ~40% of the rheMac8 *pro-pol* sequence before this breakpoint is no more similar to the other two sequences than it is to a range of other PcEV loci, while the sequences of all three downstream are very similar. We can show this phylogenetically (Fig. S1), where a tree of *pro-pol* sequences before the breakpoint shows that the rheMac8 sequence is not recovered in the same (well-supported) clade as the rheMac2 and macFas5 sequences, while a tree of the *pro-pol* sequences after this break-point shows all three sequences recovered together in a well-supported clade (as you would expect and as seen in the other loci). This represents a clear assembly error: indeed, the locus in the earlier rhesus build and in the cynomolgus macaque are in the antisense direction while the locus in the later rhesus build is in the sense direction. Consistent with widespread assembly problems, only around half of the 'intact' loci in the rhesus genome are intact in the cynomolgus genome, and *vice versa*. We also see much more fragmentation in PcEV than in the human HERVK loci (Subramanian et al., 2011), which belong to an ERV lineage that is older than PcEV (Belshaw et al., 2004). The sequencing of ERVs in the human genome benefited from them being put

within individual BACs (Bacterial Artificial Chromosomes) (Lander et al., 2001), which avoided the problem of trying to assemble simultaneously the sequences of multiple, but very similar, ERV loci from short NGS reads.

We thus fear that these assembly problems may be obscuring the presence of more recently integrated, and hence potentially active, loci. The intact loci we have recovered are all several millions of years old (Figure 25), with the most recently integrated being locus chr9:50301699-10531 discussed above. The two LTRs of this locus differ by two substitutions that are common to both builds (substitutions not in both builds are treated as sequencing errors). Given an estimated rate of nucleotide substitution of ~$1x10^{-9}$/nucleotide/year (Magiorkinis et al., 2015) the presence of two substitutions within the 512nt long LTRs suggests an age of ~1.9 million years (see Methods in section 7.27). Such loci may therefore have accrued inactivating missense substitutions even in the absence of nonsense substitutions or frameshifting indels. We found one locus that had identical LTRs, and which also appeared to be heterozygous (as expected from a young locus), chr1:55452680-60247, but this locus was fragmentary and also clearly misassembled as with chr9:50301699-10531 above (detailed in supplementary information in section 12).

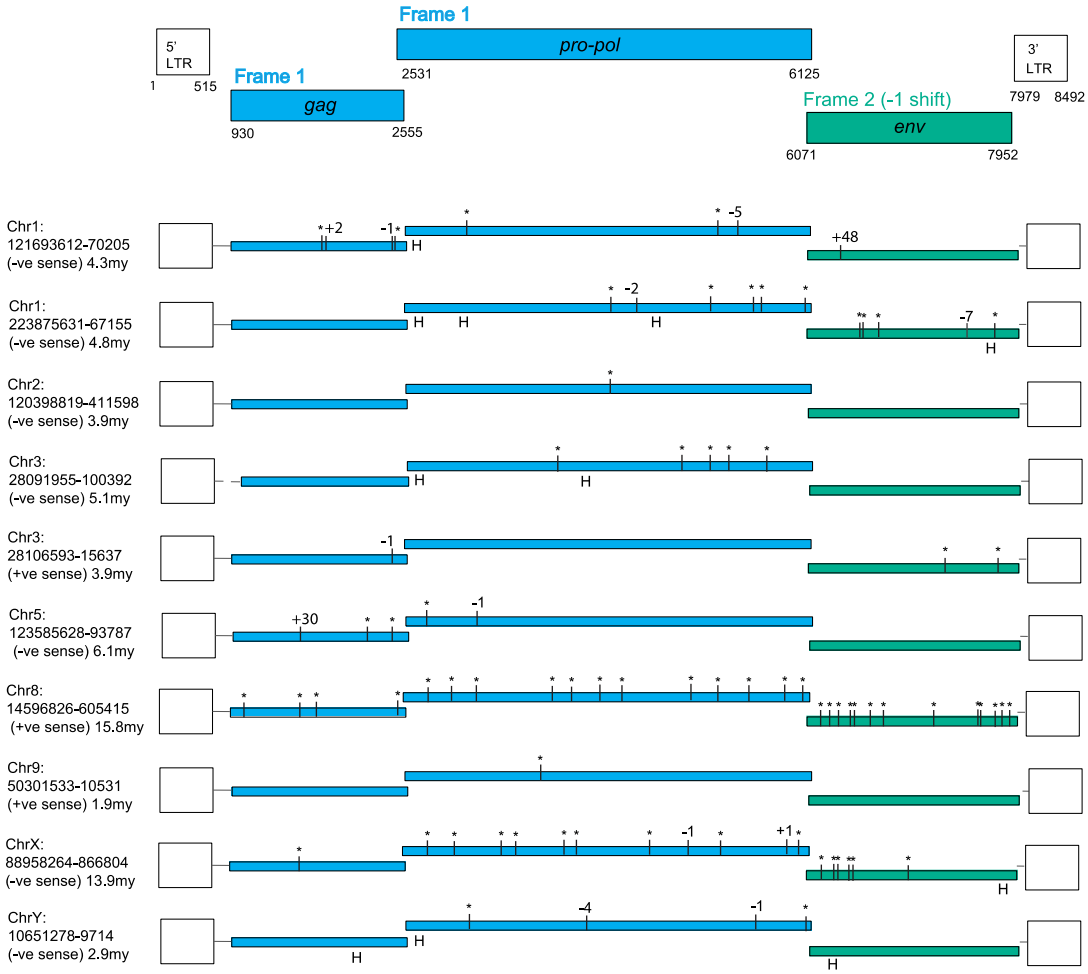


Figure 25 : Bioinformatic analysis suggests PcEV protein production but not replication.

PcEV loci for which we have complete sequences in the rhesus macaque genome with all interruptions to the reading frames indicated. Premature stop codons are shown as asterisks and frame-shifting (or large in-frame) indels are shown by the number of nucleotides involved ("H" marks an indel within a homopolymer, which we assume are likely to be sequencing errors). Results are from the rheMac8 build with all interruptions either confirmed or corrected using the earlier rheMac2 build from the same animal, i.e. interruptions were treated as sequencing errors if they are present only in one build (exceptions are the loci on the X and Y chromosomes – see below). Details of any assembly problems in these loci are in SI, which also contains multiple alignments for the three genes, the LTR alignments used for dating, the flanking regions used to determine locus homology across genomes, and reference (consensus) sequences both for the complete provirus and for individual genes. Our *pro-pol* alignment starts from the position suggested in Mang et al. (Mang et al., 1999). The reference coordinates here include a one nucleotide gap inserted at the end of *pro-pol* to incorporate the frameshift in *env*. Unusually for a retrovirus, *gag* and *pro-pol* are translated in the same frame with suppression of the *gag* stop codon (Kato et al., 1987).

8.4.3 PcEV is transcriptionally active in multiple lymphoid tissues of the macaque.

Next, we sought to determine levels of detectable cellular PcEV transcripts taken mostly at *post-mortem* by isolating RNA from PBMCs (Peripheral blood Mononuclear Cells), spleen, thymus, PLN (Peripheral Lymph Node), and MLN (Mesenteric Lymph Node). Co-amplification of GAPDH with PcEV provided a measure of cell-associated PcEV transcriptional activity in each cell preparation (Figure 26). From a total of 36 tissues preparations analysed, we determined a mean level of 36 copies PcEV RNA /1000 copies of GAPDH (range = 6-112). PcEV therefore appears to be effectively transcribed in tissues, with transcription levels lowest in the thymus and MLN (in both, mean = 25 PcEV copies per 1000 copies GAPDH). Values were higher in the Spleen (mean = 43) and PBMCs (mean = 47), and highest in the PLN (mean = 71). However, these differences of expression across tissues are not statistically significant (Kruskal-Walis test, p-value=0.08).

Interestingly, the level of tissue PcEV transcription was not significantly higher in infected compared to uninfected individuals (Figure 27A), although this may reflect the small sample size. In addition, within infected animal there is not a significant relationship between both cellular and plasma levels of SIV and PcEV (Spearman test: cellular SIV vs PcEV, r=0.14, p-value=0.46; plasma SIV vs PcEV, r=-0.17, p-value=0.41) (Figure 27B and C). Overall, our observations show PcEV to be transcriptionally active in all tissues tested, with SIV infection having no influence.

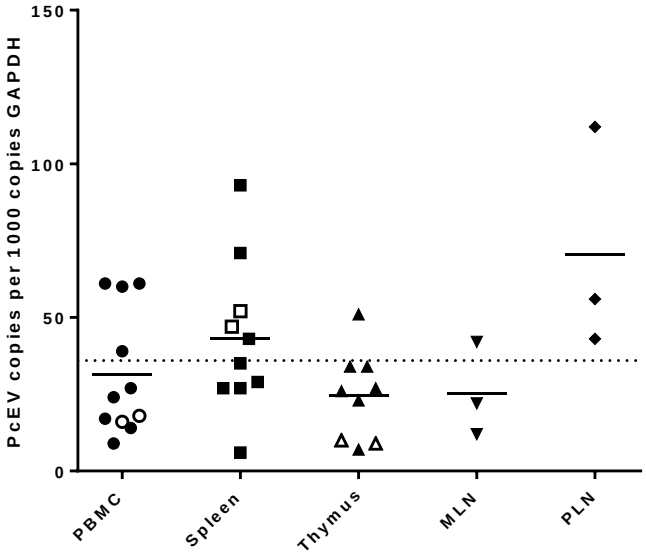


Figure 26 : PcEV is transcriptionally active in multiple lymphoid tissues of the macaque.

RNA was extracted from the tissues of non-infected or infected animals with one of 4 different SIV strains, followed by DNAse treatment. The absence of DNA contaminant was confirmed by RT-qPCR without RT enzyme. In each group, each open symbol denotes PcEV RNA copies from a single uninfected individual and each closed symbol denotes PcEV RNA copies from a single SIV-positive individual. Means for each tissue are indicated as horizontal bars and include PcEV copy number from infected & uninfected tissues. The dashed line represents the mean for all tissues (~36 copies per 1000 copies GAPDH).

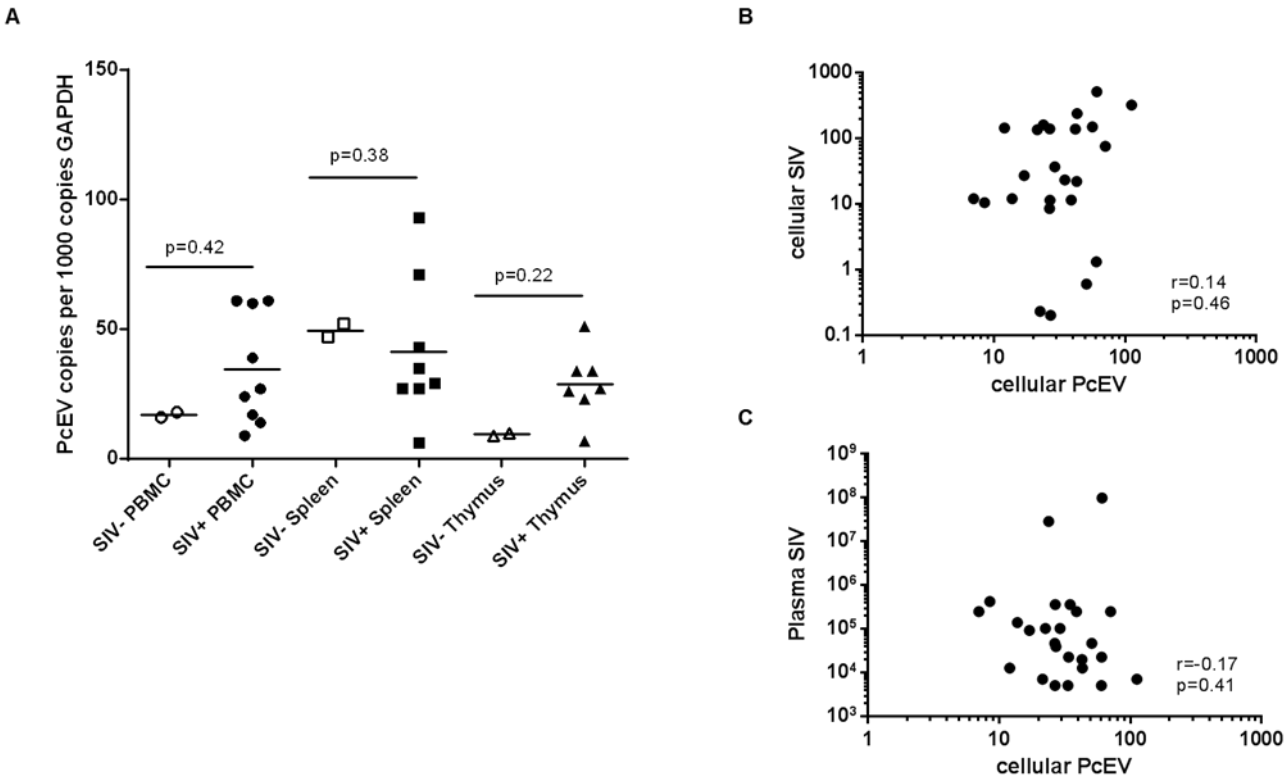


Figure 27 : SIV infection does not affect significantly PcEV expression.

(A) Each open symbol denotes PcEV RNA copies from a single uninfected individual, and each closed symbol denotes PcEV RNA copies from a single SIV-positive individual. This figure shows data in figure 26, but with SIV-infected and non-infected individuals per tissue type. P-values from a Mann-Whitney comparison of PcEV values from SIV- and SIV+ individuals is shown for each tissue. (B-C) Spearman correlation tests with coefficient and p-values indicated on each graph. Within infected animal, there is not a significant relationship between both cellular and plasma levels of SIV and PcEV.

8.4.4 Low levels of PcEV RNA present in the plasma of some infected individuals.

Assuming that free retroviral RNA transcripts in plasma must come from circulating viral particles (Bhardwaj et al., 2014; Karamitros et al., 2016), we measured PcEV RNA levels by RT-qPCR in macaque plasmas previously challenged with four different strains of SIV at or around the time of peak SIV viremia. These were SIVmac239 (n =5), SIVmac251 (n=5), SIVmacC8 (n= 6) and SIVsmE660 (n=6) with 5 plasmas from non-infected animals as controls. SIVmacC8 is an attenuated nef-disrupted variant of wild-type SIVmac251/32H (Jiang et al., 2016) As shown in figure 28A, only 7/27 plasmas exhibited detectable levels of PcEV (39; 84; 94; 139; 70; 35; 165 copies per ml plasma) and these 7 were distributed across all 4 SIV strains. PcEV was undetectable in all 5 control plasmas. All RNA preparations from plasma were subjected to DNAse

treatment and were negative for the RT- reaction, which controls for the potential co-amplification of contaminating genomic DNA, hence the PcEV molecules detected were definitely from an RNA template (see supplemental table S1 in supplementary information). However, PcEV levels in the 7 plasmas were low, in comparison with the profoundly viremic SIVs, the lowest SIV level was > 10^3 (5.14 x 10^3 SIV RNA copies/ml).

As PcEV RNA was only detected in SIV-infected individuals, we assessed whether SIV infection could explain the production, even at low level, of PcEV in the plasma. However, when the presence or absence of PcEV RNA in the plasma of individuals was compared between pooled data from non-infected and SIV-infected individuals, no significant difference was found (Figure 28B). Samples before and after infection were available for 3 individuals but PcEV RNA could not be detected in those individuals either before or after SIV challenge (E79, E80, E81, see supplemental table S1 in supplementary information).

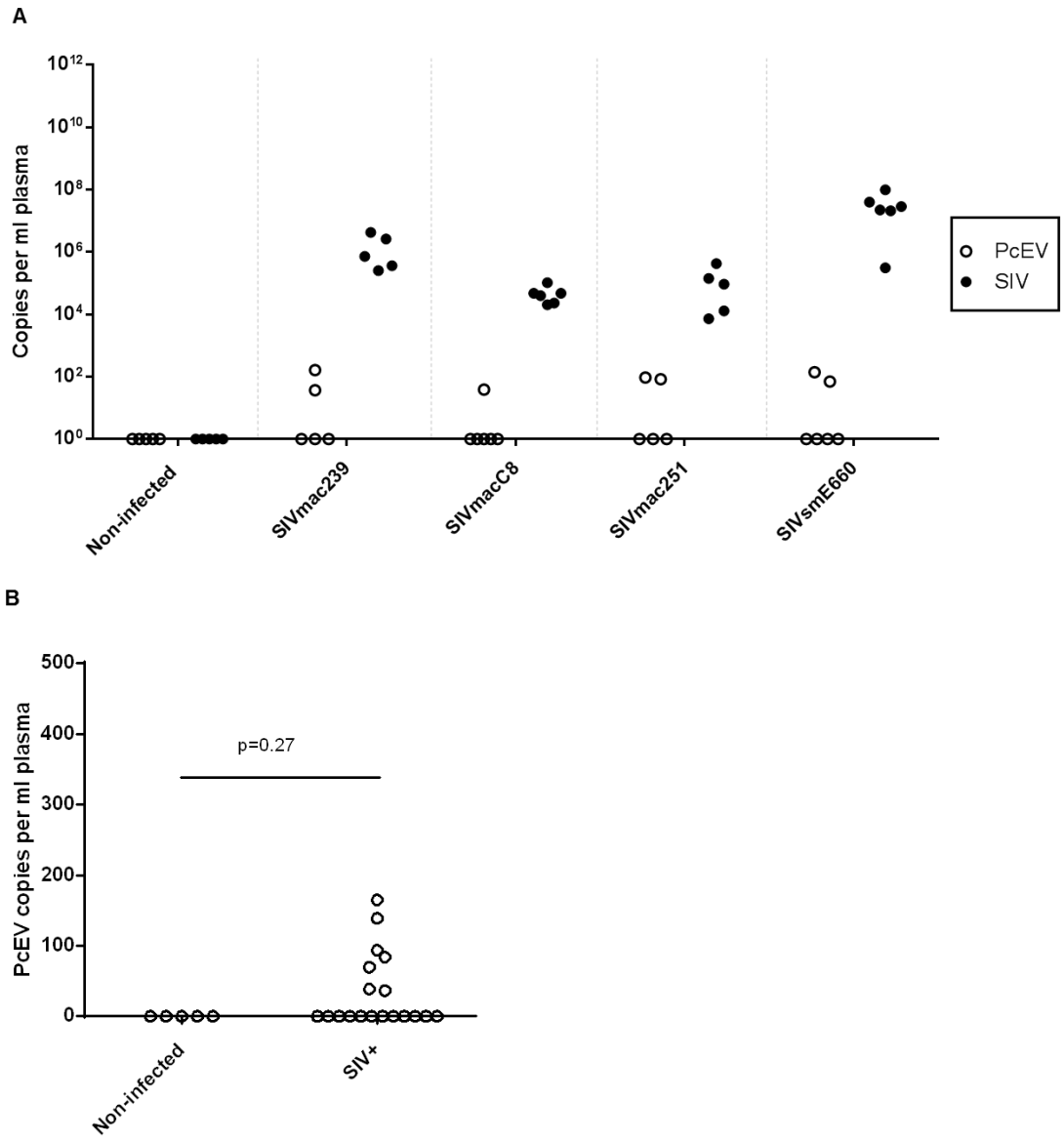


Figure 28 : Low levels of PcEV RNA present in the plasma of some infected individuals.

(A) RNA was extracted from the plasmas of animals, non-infected or infected with one of four different SIV strains (SIVmac239, SIVmacC8, SIVmac251 and SIVsmE660) followed by DNAse treatment. The absence of DNA contamination was confirmed in all experiments by simultaneous RT-qPCR without the RT enzyme. Only results from RT+ reactions are shown in the graph. The same SIV standards were used for all SIVs qPCR. Open circles denote PcEV RNA copies, and closed circles denote SIV RNA copies, from a single individual. (B) Statistical comparison of presence versus absence of PcEV in the plasma, pooling the data from non-infected individuals and SIV-infected individuals. Each open circle denotes PcEV RNA copies from a single individual. Using a Fisher's Exact Test, no significant difference was found.

8.4.5 STAT1 transcription levels correlate strongly with cell-associated PcEV but not with plasma or cell-associated SIV.

To determine any relationship between PcEV expression and IFN-1 responses after SIV challenge we measured STAT1 by RT-qPCR in all tissue samples and plotted against PcEV copies in the same tissues. A robust non-parametric test showed PcEV and STAT1 to be significantly correlated (Spearman Test, r=0.47; p=0.0091). Logarithmic transformation (base 10) of PcEV and STAT1 values produced distributions not significantly different from normal, so we also used a Pearson Test to measure the strength of the association. This showed a strong positive correlation between PcEV and STAT1 copy number per 1000 copies GAPDH (Pearson factor, r=0.48, p-value = 0.0075) (Figure 29).

In contrast, as also shown in figure 29, STAT1 levels were not strongly correlated to either cellular or plasma SIV levels in infected animals. We carried out a Spearman test which showed no relationship with STAT1 (Spearman Test: cellular SIV vs STAT1, r=0.13; p-value=0.49; plasma SIV vs STAT1, r=0.25; p-value=0.22).

It is to be noted that, as expected, STAT1 is upregulated by SIV infection compared to non-infected individuals (Figure 30). PBMCs from infected individuals showed cellular levels of STAT1 that were higher than that in naïve individuals. The small sample sizes prevent extrapolation from the non-significant results in the other tissues, but a previous study showed a drastic increase in STAT1 levels in the blood only (Ferguson et al., 2014).

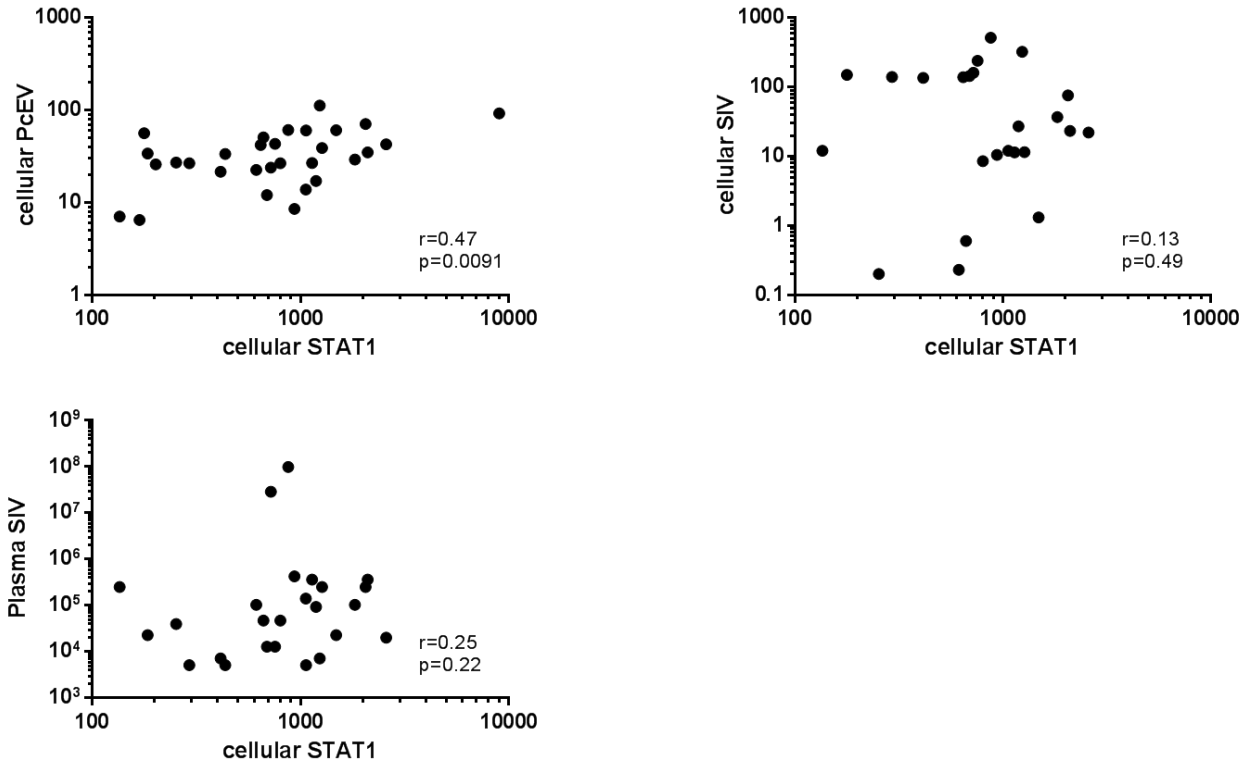


Figure 29 : STAT1 transcription levels correlate strongly with cell-associated PcEV but not plasma or cell-associated SIV.

Spearman correlation tests with coefficient and p-values indicated on each graph. Pearson correlation test after logarithmic transformation of cellular PcEV and STAT1 revealed the following: Log PcEV vs Log STAT1, r=0.48, p-value=0.0075.

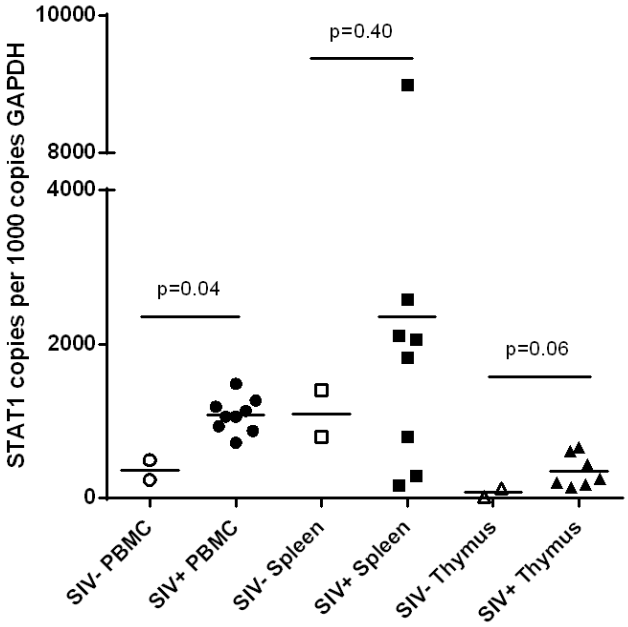


Figure 30 : STAT1 is significantly increased in the PBMCs of SIV-infected animals.

Each open symbol denotes PcEV RNA copies from a single uninfected individual, and each closed symbol denotes PcEV RNA copies from a single SIV-positive individual. Means for each SIV-infected or non-infected tissue are indicated as horizontal bars with p-values given for a Mann-Whitney comparison of the mean expression level in SIV- and SIV+ individuals for each tissue.

8.4.6 At least eight examined PcEV loci are likely to be currently transcribing.

As the in vivo analyses appeared to strongly infer that PcEV is transcriptionally active at some level, we sought to understand further which loci found in the genome may be implicated in PcEV transcription. Hence, we predicted the location of general transcription features, such as TATA box, CCAAT boxes, and direct repeat (DR) regions based on previous reports (see next paragraph), and transcription factors specific for those regions. Based on whether the core of main landing sites is disrupted, we inferred whether or not the individual PcEV locus was likely to be transcribed. Also, since there we found an association between PcEV and STAT1 transcription levels, we also sought to identify transcription factors linked to inflammation which could transactivate the PcEV LTR during infection, potentially explaining the link between PcEV and an ISG. For this purpose we also included some transcription factors suggested to induce ERVs like NF-kB members during inflammation (Uleri et al.,

2014), or activated during HIV infection such as the STAT1 (Ferguson et al., 2014) and STAT3 (Fan et al., 2015) (see methods in section 7.28).

A consensus of PcEV LTR sequence was built based on an alignment LTR sequences from eight loci analysed (Figure 25), excluding only the two very old loci. The LTR consensus was submit as a query on ALGGEN-PROMO online tool (see methods in section 7.28). We then searched for a list of transcription factors selected based on studies in *Gammaretroviruses*, namely MuLV (Speck and Baltimore, 1987), PcEV (Mang et al., 1999) and HERV-W. The list is enumerated in the main Materials and Methods section. Only primate-specific binding sites were searched for. From the list submitted, only sites for TBP [T00794]; GR-beta [T01920]; GR-alpha [T00337]; C/EBPbeta [T00581]; YY1 [T00915]; NF-1 [T00539]; C/EBPalpha [T00105]; NF-Y [T00150]; STAT3 [T01493]; GATA-2 [T00308]; GATA-1 [T00306] were found. TBP revealed the position of TATA box at position nt 349 on the LTR consensus (382 on Figure 31) which correlated with the one described by Mang et al. Since TATA box usually has a minimal promoting activity (Lee, 2003), loci without disruption in the core sequence of TBP binding sequence ((A/G)(A/G)(G/T)(A/G)TATAAA) was considered to take part to transcription. The core sequence of TBP binding site (TATAAA) is not disrupted in all 8 loci. Another TBP was predicted at position 34 (51 on Figure 31), but is unlikely real, since it is not well conserved among loci and in other *Gammaretroviruses*, TATA box is not found in that position (HERV-W, MuLV) (Lee, 2003; Luciw and Leung, 1992). Transcription factors such as GR-alpha, GR-beta and C/EBPbeta from ALGGEN-PROMO possess a very short core sequence (only 3 nucleotides; see methods in section 7.28) and so were predicted to be present in high frequency in the sequence. Enhancers usually harbor binding sites for many transcription factors (Speck and Baltimore, 1987). We therefore focused on regions that were predicted to bind many transcription factors. Interestingly, the depicted regions corresponded with description by Mang et al. (1999), and so were named in accordance with that study. We found GR-alpha, GATA1 and GATA2 sharing binding regions (DR1, DR2, DR3); GR-beta, C/EBPbeta, NF-Y, sharing CCAAT box-associated binding regions (CCAAT box 1, 2, and 3). One CCAAT box region had up to 4 binding sites, including GR-beta, C/EBPbeta, NF-Y and C/EBPalpha (CCAAT box 4). Interestingly, such CCAAT box directly upstream TATA box was shown to efficiently promote transcription activity of HERV-W LTR (Lee, 2003). As a result of this finding,

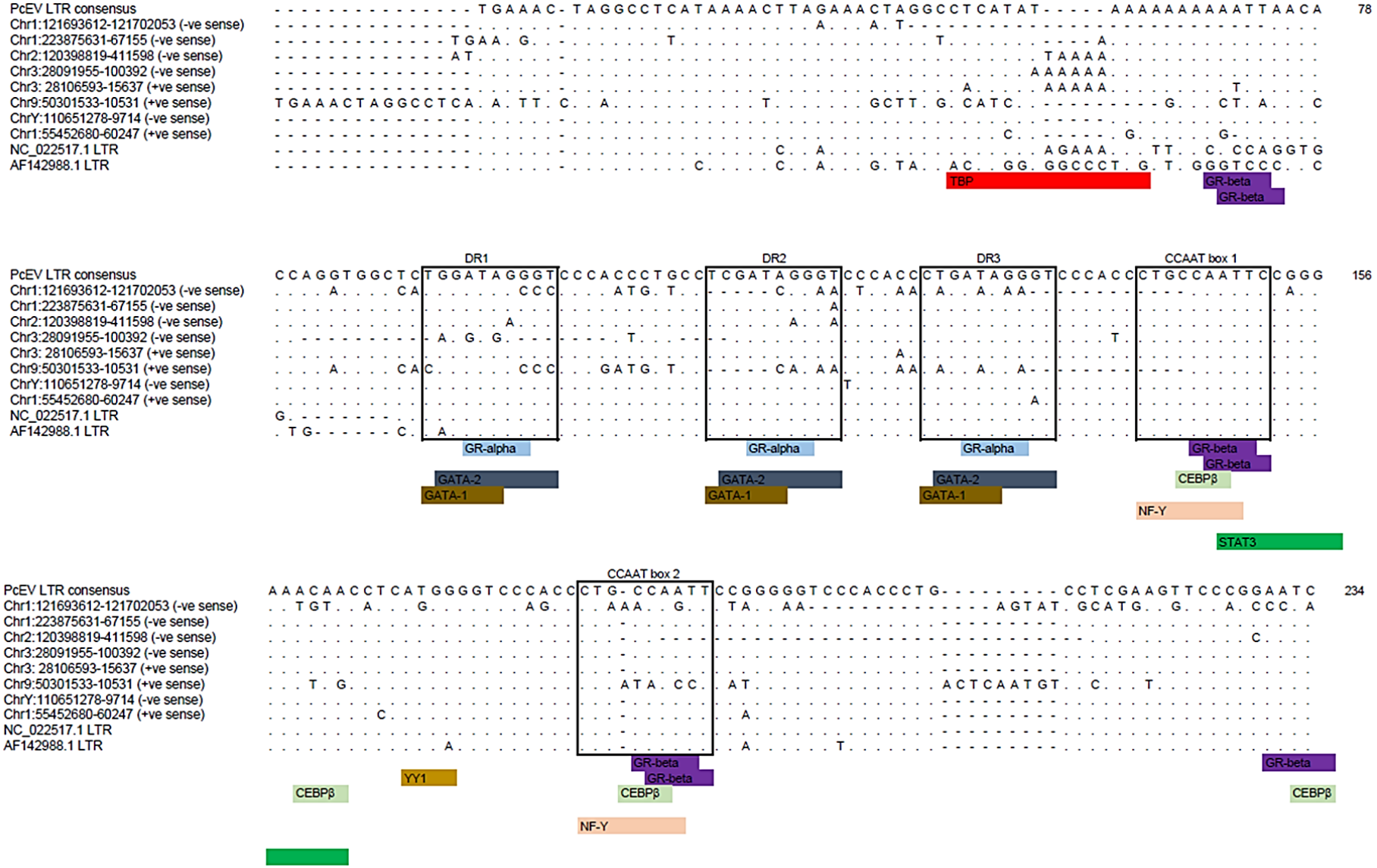
we considered that loci without mutation in the core sequence CCAAT efficiently transcribe (a "+" per non-disrupted CCAAT in table 4). Finally, in the DR regions, we considered the GATAGGG as core sequence (compilation of GR-alpha, GATA1 and GATA2 core taken together; a "+" per non-disrupted GATAGGG in table 4). The table 4 summarized our analysis.

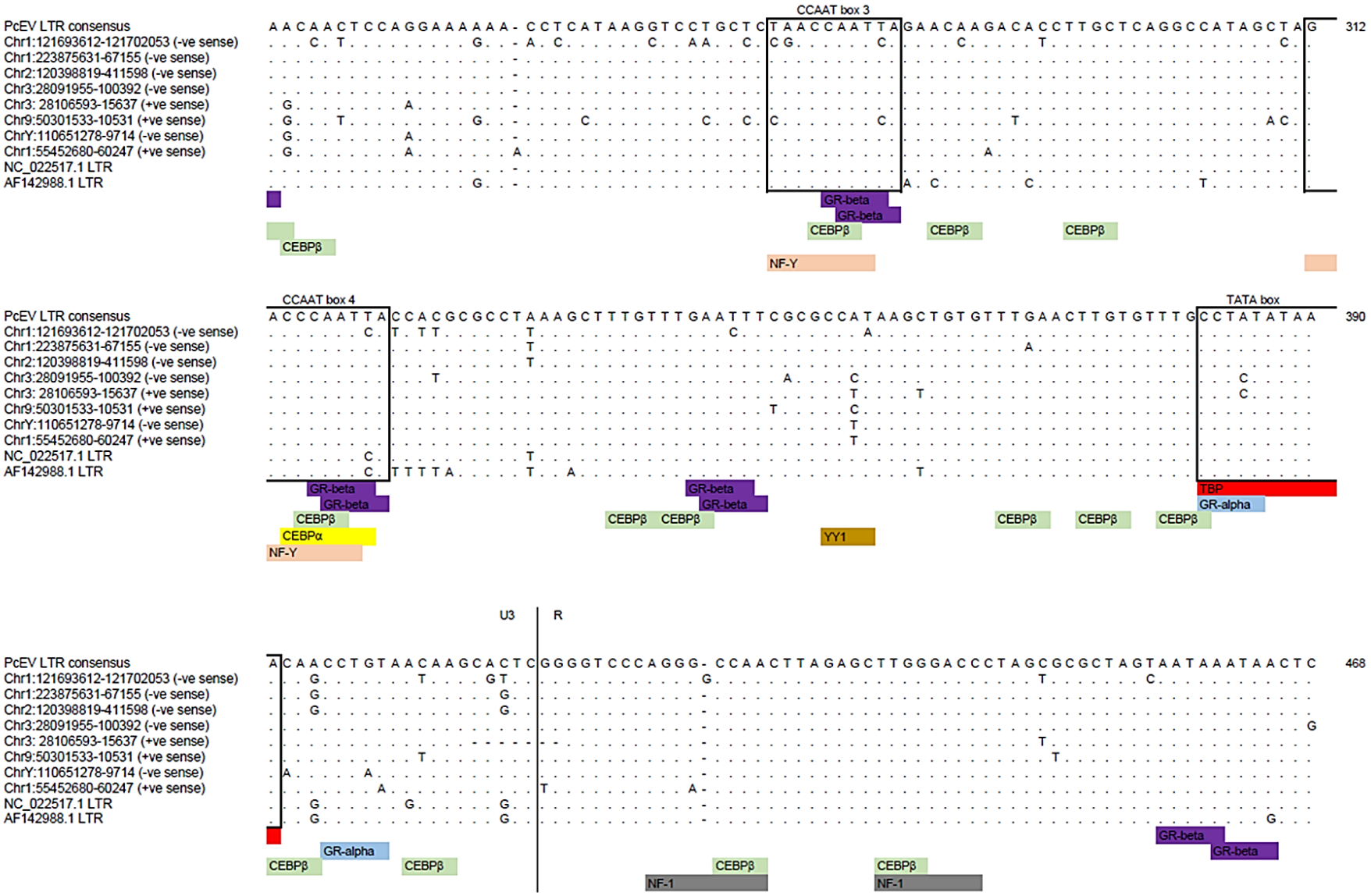
Regarding inflammation-related transcription factors, only STAT3 at position 127 (149 on Figure 31) was predicted and thus we cannot easily explain our observed relationship between STAT1 and PcEV by STAT1 upregulating PcEV during inflammation/infection.

Table 4 : Summary of transcription factor binding domains

All 8 loci of PcEV are likely actively transcribed. Some of them harbor many enhancers and so are suspect to contribute more to transcription. Yes: no mutation in the TATA box core (TATAAA); + or -: no mutation or at least one mutation, in the DR1 core (GATAGGG), CCAAT box core (CCAAT).

	TATAAA	DR1	DR2	DR3	CCAAT box 1	CCAAT box 2	CCAAT box 3	CCAAT box 4	Potential strengh
Chr1:121693612-121702053 (-ve sense)	Yes	-	-	-	-	-	+	+	++
Chr1:223875631-67155 (-ve sense)	Yes	+	+	+	+	+	+	+	+++++++
Chr2:120398819-411598 (-ve sense)	Yes	-	-	+	+	-	+	+	++++
Chr3:28091955-100392 (-ve sense)	Yes	-	+	+	+	+	+	+	++++++
Chr3: 28106593-15637 (+ve sense)	Yes	+	+	+	+	+	+	+	+++++++
Chr9:50301533-10531 (+ve sense)	Yes	-	-	-	-	-	+	+	++
ChrY:110651278-9714 (-ve sense)	Yes	+	+	+	+	+	+	+	+++++++
Chr1:55452680-60247 (+ve sense)	Yes	+	+	-	+	+	+	+	++++++


PcEV LTR consensus
Chr1:121693612-121702053 (-ve sense)
Chr1:223875631-67155 (-ve sense)
Chr2:120398819-411598 (-ve sense)
Chr3:28091955-100392 (-ve sense)
Chr3: 28106593-15637 (+ve sense)
Chr9:50301533-10531 (+ve sense)
ChrY:110651278-9714 (-ve sense)
Chr1:55452680-60247 (+ve sense)
NC_022517.1 LTR
AF142988.1 LTR
78
TBP
GR-beta
GR-beta
DR1
DR2
DR3
CCAAT box 1
156
GR-alpha
GR-alpha
GR-alpha
GR-beta
GR-beta
GATA-2
GATA-1
GATA-2
GATA-1
GATA-2
GATA-1
CEBPβ
NF-Y
STAT3
CCAAT box 2
234
GR-beta
GR-beta
GR-beta
YY1
CEBPβ
CEBPβ
CEBPβ
NF-Y


CCAAT box 3
PcEV LTR consensus
Chr1:121693612-121702053 (-ve sense)
Chr1:223875631-67155 (-ve sense)
Chr2:120398819-411598 (-ve sense)
Chr3:28091955-100392 (-ve sense)
Chr3: 28106593-15637 (+ve sense)
Chr9:50301533-10531 (+ve sense)
ChrY:110651278-9714 (-ve sense)
Chr1:55452680-60247 (+ve sense)
NC_022517.1 LTR
AF142988.1 LTR
312
GR-beta
GR-beta
CEBPβ
CEBPβ
CEBPβ
CEBPβ
NF-Y
CCAAT box 4
TATA box
390
GR-beta
GR-beta
GR-beta
GR-beta
TBP
GR-alpha
CEBPβ
CEBPβ
CEBPβ
CEBPβ
CEBPβ
CEBPβ
CEBPα
YY1
NF-Y
U3
R
468
GR-alpha
GR-beta
GR-beta
CEBPβ
CEBPβ
CEBPβ
CEBPβ
NF-1
NF-1

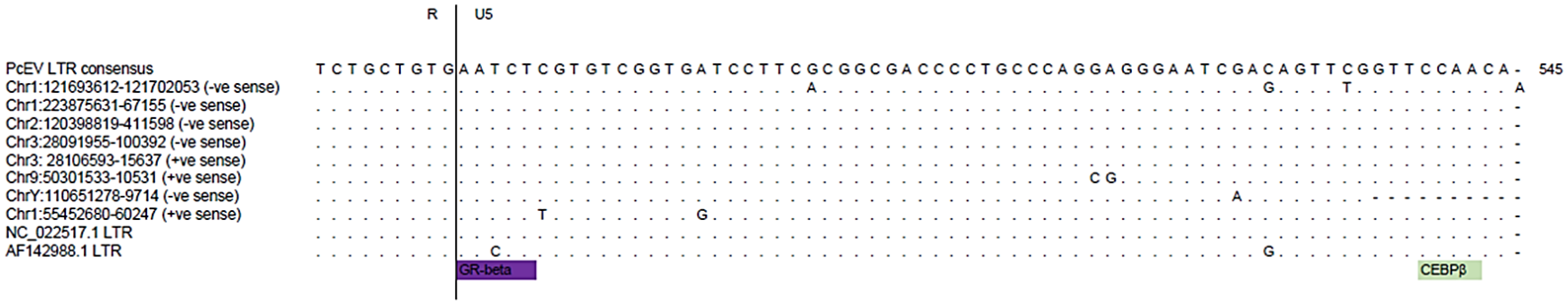


Figure 31 : Prediction of transcription factors binding sites along with enhancers region and TATA box, in LTRs from 8 PcEV loci found in macaque genome.

LTRs from NC_022517.1 and AF142988 was added to the analysis.

8.4.7 Discussion

To investigate a possible role for ERVs in innate signalling, we exploited the SIV/macaque model where the timing and nature of the exogenous infecting agent is well-defined and controlled for, and complements insights from the mouse and human. We find good evidence for PcEV RNA at low levels in the plasma of some infected macaques. The positive relationship between PcEV expression and STAT1 transcript levels – much stronger than the relationship between SIV and STAT1 – is highly suggestive of a relationship between ERV transcriptional activity, at the cellular level, and induction of the interferon response.

Modelling the role of ERVs in non-human primate models is dependent on defining the relevant ERV species in the non-human primate species studied. How comparable are our findings to what may be happening in humans? Earlier studies reported the presence of HERVK (HML2) RNA in human plasma from HIV patients (Contreras-Galindo et al., 2012) but more recent studies have failed to replicate this (Bhardwaj et al., 2014; Karamitros et al., 2016). We have yet to resolve whether PcEV RNA in the plasma results from replicationally active (infective) loci.

Although the macaque may differ from humans in possibly having – as yet unrecovered – replication-competent PcEV loci, the level of cell-associated transcription is very similar to that found in the most active human ERV lineage, HERV-K (HML2): namely ~30 copies per 1000 copies GAPDH in the PBMCs of non-acute HIV infection cases, and ~15 RNA copies in PBMCs from uninfected individuals (Bhardwaj et al., 2014).

There are examples of human ERVs for which the expression is enhanced in presence of IFN or viral infection. HERV-K18 env expression is increased after IFN treatment *in vitro* (Stauffer et al., 2001). HERV-K (HML2) is increased in PBMCs from HIV-1-infected individuals (Bhardwaj et al., 2014; Contreras-Galindo et al., 2012, 2007). The expression of HERV-W, a gammaretrovirus, is increased by influenza infection *in vitro* (F. Li et al., 2014; Nellåker et al., 2006).

Not only human ERVs seem to respond to viral insults, some were suggested to take part to the triggering of the immune response. Recombinant Env from HERV-W

activates immune cells *in vitro* through TLR4 (Rolland et al., 2006). In addition, HERV-K18 (Stauffer et al., 2001) and HERV-W (Perron et al., 2001) Envs were proposed to act as superantigens and activate T cells in a polyclonal fashion *in vitro*.

In our study, although it may be due to the sample size, SIV failed to significantly induce PcEV transcription and no correlation between levels of SIV and PcEV transcripts was found in infected individuals. Also, paired samples from individuals before and after infection were not available but would have provide a better measure of the impact of SIV on PcEV, if any. In addition, the apparent absence of STAT1 and NF-κB binding motifs on LTRs of PcEV loci analysed, means it is unlikely that those proteins are directly binding to the LTRs, then driving PcEV transcription. If it is right, this leaves the possibility of PcEV being involved in triggering the immune response.

Two main mechanisms can be imagined (Hurst and Magiorkinis, 2015). The first implies binding of ERV particles to innate receptors such as TLRs (the exogenous route). The second implies potential conversion of ERV cytoplasmic transcripts into cDNA that trigger innate sensors linked to the RIG-I–cGAS pathways (the endogenous route). In our study, we could not find PcEV transcripts in the sera from all individuals, suggesting PcEV is unlikely to produce a relevant number of virions. In this scenario, we would not expect PcEV to use the exogenous route to trigger innate receptors. Nevertheless, PcEV might perhaps trigger TLRs through released viral components, such as Env, as it is the case for HERV-W for which Env was shown to activate immune cells *in vitro* through TLR4 (Rolland et al., 2006). Antibodies responses to such Env are found elevated in sera from MS patients, as compared to healthy controls (reviewed in Christensen (2016)). In this case, it could be interesting to look for PcEV structural proteins into sera, using ELISA.

Regarding the endogenous route, we found that PcEV actively transcribes in primates. Endogenous retroviral RT was suggested to reverse transcribe viral RNA intracellularly in Human (Contreras-Galindo et al., 2017; Dube et al., 2014). In autoimmune Trex-1-deficient mouse, ERVs DNA accumulate in the cytoplasm and are associated with inflammatory phenotypes (Stetson et al., 2008); which can be treated using reverse transcriptase inhibitors (Beck-Engeser et al., 2011). In addition, mouse ERVs were

shown to bind to the RNA sensor RIG-I (Zeng et al., 2014). So, in primate, the remaining caveat concern ERV binding to innate sensors; but it is highly plausible.

Finding RNA in the plasma suggests that viral particles are present in the cell-free plasma of some but not all individuals. Although we did not find any potentially replication-competent PcEV in the reference macaque genome sequences, this may reflect a combination of (a) genome assembly problems and (b) the likelihood that such loci would be unfixed, i.e. present only in the genomes of some individual macaques (Marchi et al., 2014). We therefore speculate that macaques with very low levels of circulating PcEV may harbor a replication-competent PcEV locus responsible for the RNA found in the plasma. Alternatively, it has been shown in the mouse model that recombination between replication-defective loci can reconstitute replication-competent loci with resulting infective virions (Young et al., 2012).

Regarding a previous report in the baboon (Mang et al., 1999), the relatively low presence of PcEV viral particles could have been expected as PcEV 3'LTR lacks a DR1 repeat that is critical for viral packaging. However this way of viral packaging concerns *Alpharetroviruses* (Aschoff et al., 1999), and we are aware that *Gammaretroviruses* usually harbor a packaging signal Ψ involved in viral packaging, upstream of gag ORF (Maetzig et al., 2011).

Our observation is purely correlative. Whether ERV are directly implicated in innate response to an exogenous viral infection in primates needs to be assessed: (a) by looking for the binding of ERV nucleic acids to innate sensors (Zeng et al., 2014), (b) by knock-down of ERV transcription followed by a measurement of the innate response.

In addition, we could also analyse more retroviral lineages, namely the other two recently active ERVs, CERV and SERV (Magiorkinis et al., 2015). Our preliminary analysis suggests they are similar: we find single examples in each of single loci with full-length ORFs in both lineages. Also, we could expand the bioinformatic analysis to look for novel loci in the other macaque genomes that are available as unassembled datasets.

We could also look at LINEs, which unlike ERVs are known to be copying in the human population and which have potentially replication-competent loci in the rhesus macaque reference sequence (Han et al., 2007). Considering both ERVs and LINEs together may be more relevant and provide a wider angle to answer the question of whether retroelements are involved in innate response.

We suggest that the current study suggests a role for ERVs in line with the model proposed by Volkman and Stetson, and that this could be tested by further study. Even if ERVs do not play a role but are responsive to ISG induction, this requires clarification.

9 GENERAL DISCUSSION & PERSPECTIVES

9.1 Tissues and tumours permissive to HERV-K (HML2) expression

9.1.1 HERV-K (HML2) expression in 'healthy' tissues

There are studies reporting HERV-K (HML2) expression in human tissues in healthy conditions. The general picture is that HERV-K (HML2) expression seems to be limited in all healthy tissues, except from those that are permissive: mainly testis, placenta, prostate, embryonic tissues (Kämmerer et al., 2011; Pérot et al., 2012). To better define permissive tissues for HERV-K (HML2), standard curve-associated qPCR or RNAseq should be used, so that we could more accurately define the levels of expression per tissue. Also, not every healthy tissue is available. If such challenges are addressed accurately, it could give clues on potential involvements of those retroelements in 'normal' state. In fact, it could be easier to determine, or speculate on a role when a tissue-specific expression has been assessed. Even though there are many studies checking for expression, it is hard today to draw a list of tissues that express HERV-K (HML2) in their healthy conditions. In my thesis, even if we could check for HERV-K (HML2) expression in the nerves, we did not have a tool to accurately measure the levels of HERV-K (HML2) via qPCR. I observed that HERV-K (HML2) proteins can be found in the nerves. There were more nerves without HERV-K (HML2) staining, however the amount of nerves that were positive to the staining was non- negligible. Also, we expect a polymorphism of HERV-K (HML2) loci among individuals (Marchi et al., 2014), it means that some people that are screened could harbor an active locus, which is still fully capable of producing HERV-K (HML2) proteins that results in a positive staining.

Another use of an exhaustive analysis of tissues that express HERV-K (HML2) in a healthy state, is an appreciation of tissues that may be targeted by side effects of HERV-K-specific therapy. In fact, if HERV-K (HML2) appear to be highly expressed in a tissue, at comparable levels than in targeted tumours, we can suspect the apparition of secondary effects, especially if the monoclonal antibody induces cell death. It is then imperative to test the effects of monoclonal antibody in healthy tissue. In our case, the

use of commercial antibody did not significantly affect the proliferation of primary Schwann cells in culture.

9.1.2 HERV-K (HML2) expression in tumours

Regarding tumours, the list of HERV-K (HML2)-positive tumours is long. It is understandable, knowing that tumours undergo demethylation and HERV-K (HML2) expression, like other mammals' ERVs, is inhibit by methylation (Depil et al., 2002; Götzinger et al., 1996; Ma et al., 2011; Niwa and Sugahara, 1981). However, the list is not exhaustive, as there are still tumours in which HERV-K (HML2) expression have not been investigated. My thesis proposes to add schwannomas to the list of HERV-K (HML2)-expressing tumours. It is interesting to enumerate how many tumours express such retroelements because novel therapeutics based on monoclonal antibodies are targeting HERV-K (HML2) Env (Wang-Johanning et al., 2012). Thus, it could be a universal anti-tumour treatment, especially for tumours that do not have any current therapeutics.

9.2 Transcription factors and motifs for PcEV and HERV-K (HML2) transcription

9.2.1 PcEV transcription

To date, only very few *in silico* study had predicted TFs on PcEV LTRs (Mang et al., 1999). Also, to my knowledge, there is not a study that experimentally tested TFs binding sites. Analogy to related *Gammaretroviruses* suggest a minimal promoter that consist of a TATA box, along with enhancer sequences that contain a CCAAT box or that consist of DRs (directed repeats). The ablation of CCAAT box motif resulted in a significant decrease of promoting activity of HERV-W, an ERV part of the *Gammaretroviruses*, suggesting that it highly promote transcription driven by the TATA box (Lee, 2003). Study of DRs in MuLV suggested a decrease of replication ability when one DR is removed (Hanecak et al., 1986; Li et al., 1987). Including our study, there are very few studies that focus on transcription factors that could drive PcEV transcription. Also, we are the first to report active transcription of PcEV in the macaque. However, we could not link TFs to transcription activity, as at the time, we did not have the tool to investigate transcription of mutated PcEV LTRs. Another point is the prediction of STAT3 as binding to PcEV LTR. We found a strong positive

association between PcEV transcription and STAT1 transcription. It is speculative to infer that PcEV transcription is driven by members of the STAT family landing on PcEV LTRs, especially when STAT3 is found instead of STAT1. However, when the prediction software is run with a higher dissimilarity margin of 15%, the same site is predicted to be bound by STAT1 and NF-κB. To assess the veracity of this, chromatin immunoprecipitation should be performed.

9.2.2 HERV-K (HML2) transcription

Regarding HERV-K (HML2), there are MITF-M, Oct1, NF-κB, NF-AT, YY1, Sp1 and Sp3 which were experimentally tested to land on LTRs out of 20 predicted (Manghera and Douville, 2013). Many more could be tested; however, it could be challenging. Another approach could require a general pull down of TFs using ERV-derived DNA baits, and try to determine candidate through mass spectrometry (Wierer and Mann, 2016). It is relevant to determine TF-associated transcription of HERV-K (HML2), as it would help understand how HERV-K is reactivated in tumours, in a detailed manner. Demethylation of HERV-K (HML2) LTRs is not the only event that drive transcription, as reports suggested that the presence of some TFs in the cell is required (Fuchs et al., 2011; St Laurent et al., 2013). In our case, we investigate the possibility that TEAD, TF at the end point of Hippo-pathway, is involved in HERV-K (HML2) transcription. It was to be tested as Hippo-pathway is upregulated in schwannoma (L. Zhou et al., 2016). It could have provided a link to HERV-K (HML2) proteins production in schwannoma. Our test revealed a significant decrease of HERV-K (HML2) *env* transcription only. But, the verteporfin seemed to produce an effect that was not only targeting YAP-TEAD complex-driven transcription. One NF-κB binding site tested by chromatin immunoprecipitation (Gonzalez-Hernandez et al., 2012) overlaps with the identified binding site for TEAD, hence it seems that it is not physically allocated to TEAD. However, it is to be noted that the use of an NF-κB inhibitor, that prevents NF-κB from translocating into the nucleus, did not show any effect on HERV-K (HML2) Env protein level in schwannoma, as assessed by western blotting (data not shown). To clear address the possibility of TEAD binding on HERV-K (HML2) LTRs, a chromatin-immunoprecipitation of TEAD followed by HERV-K (HML2) LTR-specific PCR should be performed.

9.3 Role of ERVs in diseases

ERVs are suggested to play a role in placentation, e.g. in the case of the syncytins in mammals that have been co-opted in many species (Lavialle et al., 2013). A role in gene fine-tuning has also been suggested, e.g. HERV-H LTRs appear to allow silencing of neighboring genes or transcribe long non-coding RNA involved in pluripotency (Lu et al., 2014; St Laurent et al., 2013). When a sequence retains intact ORFs, it can also be involved in interfering with new infections by the same or related viruses. Few examples are: HERV-P in primates, whose extinction is suggested to have happened with the aid of an endogenized HERV-P viral Env, such Env experimentally diminishes infection by reconstituted viral particles (Blanco-Melo et al., 2017); and enJSRV (a JSRV-related ERV) *gag* in ovins which is suggested to limit the replication of exogenous JSRV (reviewed in Arnaud et al. (2008)). More recently, the role of ERVs in immunity have been investigated. ERVs ssDNA accumulate in the cytoplasm of Trex1-deficient mice, possibly leading to autoimmune disease development (Stetson et al., 2008). In human, HERV-W is reported to stimulate inflammation through binding to TLR4 (Rolland et al., 2006), and it is suggested to be one mechanism associated with multiple sclerosis. However, very few reports investigate whether ERVs in humans or other primates are physiologically linked to the function of triggering of innate immunity. Here we propose that in the macaque, there is an association between a primate ERV lineage, PcEV, and the innate response that arises during SIV infection, as measured by STAT1 expression. However, whether the expression of ERVs are part of innate response triggering or just a consequence is still a mystery. When considering SIV infected-animals only, such association was stronger than that between STAT1 and both cellular and plasma SIV expression. Physiologically, it looks like the more ERVs are transcribed in individuals, the higher their innate response will be, as measured by an ISG; for this reason, we stressed that in primates too, ERVs could take part to the innate response process. There exist other lineages of ERVs in the macaque; also, the LINEs need to be looked at. We need to test for a wider correlation between all retroelements (mainly ERVs + LINEs) and innate response triggering. Another perspective is to unravel the causality between ERVs and innate response, meaning that ERVs in primate need to be assessed for potential binding to innate sensors, i.e. TLRs, RIG-I, NLRs.

10 CONCLUSION

The main findings of this thesis are as follow: (i) HERV-K (HML2) is overexpressed in schwannomas in comparison to Schwann cells; (ii) Such overexpression possibly involve YAP-TEAD-driven transactivation of HERV-K (HML2) transcription; (iii) Potential therapeutics for schwannomas involved HERV-K (HML2)-specific monoclonal antibodies and ritonavir, an antiretroviral; (iv) PcEV is not viraemic in macaques as not consistently retrieved in sera; (v) PcEV is actively transcribing; (vi) High levels of PcEV transcripts in cells correlate with high levels of STAT1, an ISG, validating our hypothesis (see section 6, objective 5).

11 REFERENCES

Aaronson, S.A., Hartley, J.W., Todaro, G.J., 1969. Mouse Leukemia Virus: "Spontaneous" Release by Mouse Embryo cells After Long-Term In Vitro Cultivation. PNAS 64, 87–94.

Agnihotri, S., Jalali, S., Wilson, M.R., Danesh, A., Li, M., Klironomos, G., Krieger, J.R., Mansouri, A., Khan, O., Mamatjan, Y., Landon-Brace, N., Tung, T., Dowar, M., Li, T., Bruce, J.P., Burrell, K.E., Tonge, P.D., Alamsahebpour, A., Krischek, B., Agarwalla, P.K., Bi, W.L., Dunn, I.F., Beroukhim, R., Fehlings, M.G., Bril, V., Pagnotta, S.M., Iavarone, A., Pugh, T.J., Aldape, K.D., Zadeh, G., 2016. The genomic landscape of schwannoma. Nat. Genet. 48, 1339–1348. https://doi.org/10.1038/ng.3688

Alberti, A., Murgia, C., Liu, S.-L., Mura, M., Cousens, C., Sharp, M., Miller, A.D., Palmarini, M., 2002. Envelope-Induced Cell Transformation by Ovine Betaretroviruses. J. Virol. 76, 5387–5394. https://doi.org/10.1128/JVI.76.11.5387–5394.2002

Amarzguioui, M., Prydz, H., 2004. An algorithm for selection of functional siRNA sequences. Biochem. Biophys. Res. Commun. 316, 1050–1058. https://doi.org/10.1016/J.BBRC.2004.02.157

Ammoun, S., Flaiz, C., Ristic, N., Schuldt, J., Hanemann, C.O., 2008. Dissecting and targeting the growth factor-dependent and growth factor-independent extracellular signal-regulated kinase pathway in human schwannoma. Cancer Res. 68, 5236–5245. https://doi.org/10.1158/0008-5472.CAN-07-5849

Ammoun, S., Hanemann, C.O., 2011. Emerging therapeutic targets in schwannomas and other merlin-deficient tumors. Nat. Rev. Neurol. 7, 392–399. https://doi.org/10.1038/nrneurol.2011.82

Ammoun, S., Provenzano, L., Zhou, L., Barczyk, M., Evans, K., Hilton, D.A., Hafizi, S., Hanemann, C.O., 2014. Axl/Gas6/NFκB signalling in schwannoma pathological proliferation, adhesion and survival. Oncogene 33, 336–346. https://doi.org/10.1038/onc.2012.587

Antoinette C. van der Kuyll; John T. Dekker; Jaap Goudsmit, 1995. Full-Length Proviruses of Baboon Endogenous Virus (BaEV) and Dispersed BaEV Reverse Transcriptase Retroelements in the Genome of Baboon Species. J. Virol. 69, 5917–5924.

Armbruester, V., Sauter, M., Krautkraemer, E., Meese, E., Kleiman, A., Best, B., Roemer, K., Mueller-Lantzsch, N., 2002. A novel gene from the human endogenous retrovirus K expressed in transformed cells. Clin. Cancer Res. 8, 1800–7.

Arnaud F, Varela M, Spencer TE, Palmarini M., 2008. Coevolution of endogenous betaretroviruses of sheep and their host. Cell Mol Life Sci., 65(21):3422-32. https://doi.org/10.1007/s00018-008-8500-9

Aschoff, J.M., Foster, D., Coffin, J.M., 1999. Point mutations in the avian sarcoma/leukosis virus 3' untranslated region result in a packaging defect. J. Virol. 73, 7421–9.

Astrin, S.M., Buss, E.G., Hayward, W.S., 1979. Endogenous viral genes are non-essential in the chicken [38]. Nature. https://doi.org/10.1038/282339a0

Astrin, S.M., Crittenden, L.B., Buss, E.G., 1980. Ev 2, a genetic locus containing structural genes for endogenous virus, codes for Rous-associated virus type 0 produced by line 72 chickens. J. Virol. 33, 250–5.

Astrin, S.M., Robinson, H., 1979. Gs , an Allele of Chickens for Endogenous Avian

Leukosis Viral Antigens , Segregates with ev 3 , a Genetic Locus That Contains Structural Genes for Virus. J. Virol. 31, 420–425.
Azmi, A.S., Bao, B., Sarkar, F.H., Asfar S. Azmi1, Bin Bao, and F.H.S., 2014. Exosomes in Cancer Development, Metastisis and Drug Resistance: A Comprehensive Review. Cancer Metastasis Rev. 32, 1–33. https://doi.org/10.1007/s10555-013-9441-9.Exosomes
Bader, J., 1964. The Role of Deoxyribonucleic Sarcoma Acid in the Synthesis of Rous Virus. Virology 462368.
Baltimore, D., 1970. Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature. https://doi.org/10.1038/2261209a0
Barber, R.D., Harmer, D.W., Coleman, R.A., Clark, B.J., 2005. GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. Physiol. Genomics 21, 389–395. https://doi.org/10.1152/physiolgenomics.00025.2005
Batchu, R.B., Gruzdyn, O. V., Bryant, C.S., Qazi, A.M., Kumar, S., Chamala, S., Kung, S.T., Sanka, R.S., Puttagunta, U.S., Weaver, D.W., Gruber, S.A., 2014. Ritonavir-mediated induction of apoptosis in pancreatic cancer occurs via the RB/E2F-1 and AKT pathways. Pharmaceuticals 7, 46–57. https://doi.org/10.3390/ph7010046
Beck-Engeser, G.B., Eilat, D., Wabl, M., 2011. An autoimmune disease prevented by anti-retroviral drugs. Retrovirology 8, 91. https://doi.org/10.1186/1742-4690-8-91
Bellas, R.E., Hopkins, N., Li, Y., 1993. The NF-kappa B binding site is necessary for efficient replication of simian immunodeficiency virus of macaques in primary macrophages but not in T cells in vitro. J. Virol. 67, 2908–2913.
Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Paces, J., Burt, A., Tristem, M., 2004. Long-term reinfection of the human genome by endogenous retroviruses. Proc. Natl. Acad. Sci. 101, 4894–4899. https://doi.org/10.1073/pnas.0307800101
Berkhout, B., Jebbink, M., Zsíros, J., 1999. Identification of an active reverse transcriptase enzyme encoded by a human endogenous HERV-K retrovirus. J. Virol. 73, 2365–75.
Berry, N., Ham, C., Mee, E.T., Rose, N.J., Mattiuzzo, G., Jenkins, A., Page, M., Elsley, W., Robinson, M., Smith, D., Ferguson, D., Towers, G., Almond, N., Stebbings, R., 2011. Early potent protection against heterologous SIVsmE660 challenge following live attenuated SIV vaccination in mauritian cynomolgus macaques. PLoS One 6. https://doi.org/10.1371/journal.pone.0023092
Berry, N., Stebbings, R., Ferguson, D., Ham, C., Alden, J., Brown, S., Jenkins, A., Lines, J., Duffy, L., Davis, L., Elsley, W., Page, M., Hull, R., Stott, J., Almond, N., 2008. Resistance to superinfection by a vigorously replicating, uncloned stock of simian immunodeficiency virus (SIVmac251) stimulatesreplication of a live attenuated virus vaccine (SIVmacC8). J. Gen. Virol. 89, 2240–2251. https://doi.org/10.1099/vir.0.2008/001693-0
Bhardwaj, N., Maldarelli, F., Mellors, J., Coffin, J.M., 2014. HIV-1 Infection Leads to Increased Transcription of Human Endogenous Retrovirus HERV-K (HML-2) Proviruses In Vivo but Not to Increased Virion Production. J. Virol. 88, 11108–11120. https://doi.org/10.1128/JVI.01623-14
Bhat, R.K., Rudnick, W., Antony, J.M., Maingat, F., Ellestad, K.K., Wheatley, B.M., Tönjes, R.R., Power, C., 2014. Human endogenous retrovirus-K(II) envelope induction protects neurons during HIV/AIDS. PLoS One 9, 1–13.

https://doi.org/10.1371/journal.pone.0097984
Bhatheja, K., Field, J., 2006. Schwann cells: Origins and role in axonal maintenance and regeneration. Int. J. Biochem. Cell Biol. https://doi.org/10.1016/j.biocel.2006.05.007
Blaise, S., de Parseval, N., Benit, L., Heidmann, T., 2003. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proc. Natl. Acad. Sci. 100, 13013–13018. https://doi.org/10.1073/pnas.2132646100
Blanco-Melo, D., Gifford, R.J., Bieniasz, P.D., 2017. Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors. Elife 6, 1–19. https://doi.org/10.7554/eLife.22519
Blond, J.-L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., Mandrand, B., Mallet, F., Cosset, F.-L., 2000. An Envelope Glycoprotein of the Human Endogenous Retrovirus HERV-W Is Expressed in the Human Placenta and Fuses Cells Expressing the Type D Mammalian Retrovirus Receptor. J. Virol. 74, 3321–3329. https://doi.org/10.1128/JVI.74.7.3321-3329.2000
Boller, K., Janssen, O., Schuldes, H., Tonjes, R.R., Kurth, R., 1997. Characterization of the Antibody Response Specific for the Human Endogenous Retrovirus HTDV/HERV-K. J Virol 71, 4581–4588.
Boller, K., Schonfeld, K., Lischer, S., Fischer, N., Hoffmann, A., Kurth, R., Tonjes, R. R., 2008.Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. J. Gen. Virol. 89, 567–572. https://doi.org/10.1099/vir.0.83534-0
Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., Liu, E.T., 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. 1752–1762. https://doi.org/10.1101/gr.080663.108.These
Brattås, P.L., Jönsson, M.E., Fasching, L., Nelander Wahlestedt, J., Shahsavani, M., Falk, R., Falk, A., Jern, P., Parmar, M., Jakobsson, J., 2017. TRIM28 Controls a Gene Regulatory Network Based on Endogenous Retroviruses in Human Neural Progenitor Cells. Cell Rep. 18, 1–11. https://doi.org/10.1016/j.celrep.2016.12.010
Bretscher, A., Edwards, K., Fehon, R.G., 2002. ERM proteins and merlin: Integrators at the cell cortex. Nat. Rev. Mol. Cell Biol. https://doi.org/10.1038/nrm882
Brodowska, K., Al-Moujahed, A., Marmalidou, A., Meyer zu Horste, M., Cichy, J., Miller, J.W., Gragoudas, E., Vavvas, D.G., 2014. The clinically used photosensitizer Verteporfin (VP) inhibits YAP-TEAD and human retinoblastoma cell growth invitro without light activation. Exp. Eye Res. 124, 67–73. https://doi.org/10.1016/j.exer.2014.04.011
Buscher, K., 2005. Expression of Human Endogenous Retrovirus K in Melanomas and Melanoma Cell Lines. Cancer Res. 65, 4172–4180. https://doi.org/10.1158/0008-5472.CAN-04-2983
Callahan, R., Drohan, W., Tronick, S., Schlom, J., 1982. Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. Proc. Natl. Acad. Sci. U. S. A. 79, 5503–7.
Camargo, F.D., Gokhale, S., Johnnidis, J.B., Fu, D., Bell, G.W., Jaenisch, R., Brummelkamp, T.R., 2007. YAP1 Increases Organ Size and Expands Undifferentiated Progenitor Cells. Curr. Biol. 17, 2054–2060. https://doi.org/10.1016/j.cub.2007.10.039
Chen, T., Meng, Z., Gan, Y., Wang, X., Xu, F., Gu, Y., Xu, X., Tang, J., Zhou, H.,

Zhang, X., Gan, X., Van Ness, C., Xu, G., Huang, L., Zhang, X., Fang, Y., Wu, J., Zheng, S., Jin, J., Huang, W., Xu, R., 2013. The viral oncogene Np9 acts as a critical molecular switch for co-activating β-catenin, ERK, Akt and Notch1 and promoting the growth of human leukemia stem/progenitor cells. Leukemia 27, 1469–78. https://doi.org/10.1038/leu.2013.8

Christensen, T., 2016. Human endogenous retroviruses in neurologic disease. Apmis 124, 116–126. https://doi.org/10.1111/apm.12486

Chuong, E.B., Elde, N.C., City, S.L., 2016. HHS Public Access 351, 1083–1087. https://doi.org/10.1126/science.aad5497.Regulatory

Coffin, J.M., 1992. Structure and Classification of Retroviruses, in: The Retroviridae, Vol 1. Springer US, Boston, MA, pp. 19–49. https://doi.org/10.1007/978-1-4615-3372-6_2

Contreras-Galindo, R., Kaplan, M.H., Contreras-Galindo, A.C., Gonzalez-Hernandez, M.J., Ferlenghi, I., Giusti, F., Lorenzo, E., Gitlin, S.D., Dosik, M.H., Yamamura, Y., Markovitz, D.M., 2012. Characterization of Human Endogenous Retroviral Elements in the Blood of HIV-1-Infected Individuals. J. Virol. 86, 262–276. https://doi.org/10.1128/JVI.00602-11

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

Contreras-Galindo, R., Kaplan, M.H., Leissner, P., Verjat, T., Ferlenghi, I., Bagnoli, F., Giusti, F., Dosik, M.H., Hayes, D.F., Gitlin, S.D., Markovitz, D.M., 2008. Human Endogenous Retrovirus K (HML-2) Elements in the Plasma of People with Lymphoma and Breast Cancer. J. Virol. 82, 9329–9336. https://doi.org/10.1128/JVI.00646-08

Contreras-Galindo, R., López, P., Vélez, R., Yamamura, Y., 2007. HIV-1 Infection Increases the Expression of Human Endogenous Retroviruses Type K (HERV-K) *in Vitro*. AIDS Res. Hum. Retroviruses 23, 116–122. https://doi.org/10.1089/aid.2006.0117

Contreras-Galindo, R.A., Dube, D., Fujinaga, K., Kaplan, M.H., Markovitz, D.M., 2017. Susceptibility of Human Endogenous Retrovirus Type-K to Reverse Transcriptase Inhibitors. J. Virol. JVI.01309-17. https://doi.org/10.1128/JVI.01309-17

Crawford, L.J., Walker, B., Irvine, A.E., 2011. Proteasome inhibitors in cancer therapy. J. Cell Commun. Signal. https://doi.org/10.1007/s12079-011-0121-7

Crawford, L. V, Crawford, E.M., 1961. The Properties of Rous Sarcoma Virus Purifed by Density Gradient Centrifugation. Virology 13, 227–232.

Crow, Y.J., Hayward, B.E., Parmar, R., Robins, P., Leitch, A., Ali, M., Black, D.N., van Bokhoven, H., Brunner, H.G., Hamel, B.C., Corry, P.C., Cowan, F.M., Frints, S.G., Klepper, J., Livingston, J.H., Lynch, S.A., Massey, R.F., Meritet, J.F., Michaud, J.L., Ponsot, G., Voit, T., Lebon, P., Bonthron, D.T., Jackson, A.P., Barnes, D.E., Lindahl, T., 2006. Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 cause Aicardi-Goutières syndrome at the AGS1 locus. Nat. Genet. 38, 917–920. https://doi.org/10.1038/ng1845

Danilkovitch-Miagkova, A., Duh, F.-M., Kuzmin, I., Angeloni, D., Liu, S.-L., Miller, a D., Lerman, M.I., 2003. Hyaluronidase 2 negatively regulates RON receptor tyrosine kinase and mediates transformation of epithelial cells by jaagsiekte sheep retrovirus. Proc. Natl. Acad. Sci. U. S. A. 100, 4580–5.

https://doi.org/10.1073/pnas.0837136100
de Parseval, N., Diop, G., Blaise, S., Helle, F., Vasilescu, A., Matsuda, F., Heidmann, T., 2005. Comprehensive search for intra- and inter-specific sequence polymorphisms among coding envelope genes of retroviral origin found in the human genome: Genes and pseudogenes. BMC Genomics 6, 1–11. https://doi.org/10.1186/1471-2164-6-117
Depil, S., Roche, C., Dussart, P., 2002. Expression of a human endogenous retrovirus, HERV-K, in the blood cells of leukemia patients. J. Leuk. 3, 254–259. https://doi.org/10.1038/sj/leu/2402355
Dewannieux, M., Blaise, S., Heidmann, T., 2005. Identification of a Functional Envelope Protein from the HERV-K Family of Human Endogenous Retroviruses. J. Virol. 79, 15573–15577. https://doi.org/10.1128/JVI.79.24.15573
Dewannieux, M., Harper, F., Richaud, A., Letzelter, C., Ribet, D., Pierron, G., Heidmann, T., 2006. Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. Genome Res. 16, 1548–1556. https://doi.org/10.1101/gr.5565706
Dougherty, R.M., Di Stefano, H.S., 1965. Virus particles associated with "nonproducer" Rous sarcoma cells. Virology 27, 351–9.
Dube, D., Contreras-Galindo, R., He, S., King, S.R., Gonzalez-Hernandez, M.J., Gitlin, S.D., Kaplan, M.H., Markovitz, D.M., 2014. Genomic Flexibility of Human Endogenous Retrovirus Type K. J. Virol. 88, 9673–9682. https://doi.org/10.1128/JVI.01147-14
Dupressoir, A., Marceau, G., Vernochet, C., Bénit, L., Kanellopoulos, C., Sapin, V., & Heidmann, T., 2005. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. Proc. Natl. Acad. Sci. 102(3), 725-30. https://doi.org/10.1073/pnas.0406509102
Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P., Heidmann, T., 2009. Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. Proc. Natl. Acad. Sci. 106, 12127–12132. https://doi.org/10.1073/pnas.0902925106
Dupressoir, A., Vernochet, C., Harper, F., Guegan, 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. Proc. Natl. Acad. Sci. 108, E1164–E1173. https://doi.org/10.1073/pnas.1112304108
Ehrlich, M., 2009. DNA hypomethylation in cancer cells. Epigenomics 1, 239–259. https://doi.org/10.2217/epi.09.33
Ehrlich, M., 2002. DNA methylation in cancer: Too much, but also too little. Oncogene 21, 5400–5413. https://doi.org/10.1038/sj.onc.1205651
Elbakri, A., Nelson, P.N., Abu Odeh, R.O., 2010. The state of antibody therapy. Hum. Immunol. https://doi.org/10.1016/j.humimm.2010.09.007
Escalera-Zamudio, M., Greenwood, A.D., 2016. On the classification and evolution of endogenous retrovirus: Human endogenous retroviruses may not be “human” after all. Apmis 124, 44–51. https://doi.org/10.1111/apm.12489
Esnault, C., Priet, S., Ribet, D., Vernochet, C., Bruls, T., Lavialle, C., Weissenbach, J., Heidmann, T., 2008. A placenta-specific receptor for the fusogenic, endogenous retrovirus-derived, human syncytin-2. Proc. Natl. Acad. Sci. 105, 17532–17537. https://doi.org/10.1073/pnas.0807413105

Evans, D.G.R., 2000. Neurofibromatosis type 2. J. Med. Genet. 37, 897–904. https://doi.org/10.1136/jmg.37.12.897
Fan, Y., Timani, K.A., He, J.J., 2015. STAT3 and its phosphorylation are involved in HIV-1 Tat-induced transactivation of glial fibrillary acidic protein. Curr. HIV Res. 13, 55–63. https://doi.org/10.1016/S2215-0366(16)30284-X.Epidemiology
Farré, D., Roset, R., Huerta, M., Adsuara, J.E., Roselló, L., Albà, M.M., Messeguer, X., 2003. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res. 31, 3651–3653. https://doi.org/10.1093/nar/gkg605
Fasching, L., Kapopoulou, A., Sachdeva, R., Petri, R., Jönsson, M.E., Männe, C., Turelli, P., Jern, P., Cammas, F., Trono, D., Jakobsson, J., 2015. TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. Cell Rep. 10, 20–28. https://doi.org/10.1016/j.celrep.2014.12.004
Ferguson, D., Mattiuzzo, G., Ham, C., Stebbings, R., Li, B., Rose, N.J., Mee, E.T., Smith, D., Page, M., Cranage, M.P., Almond, N., Towers, G.J., Berry, N.J., 2014. Early biodistribution and persistence of a protective live attenuated SIV vaccine elicits localised innate responses in multiple lymphoid tissues. PLoS One 9, 1–12. https://doi.org/10.1371/journal.pone.0104390
Freimanis, G., Hooley, P., Ejtehadi, H.D., Ali, H.A., Veitch, A., Rylance, P.B., Alawi, A., Axford, J., Nevill, A., Murray, P.G., Nelson, P.N., 2010. A role for human endogenous retrovirus-K (HML-2) in rheumatoid arthritis: Investigating mechanisms of pathogenesis. Clin. Exp. Immunol. 160, 340–347. https://doi.org/10.1111/j.1365-2249.2010.04110.x
Frendo, J., Olivier, D., Cheynet, V., Blond, J., Bouton, O., Vidaud, M., 2003. Direct Involvement of HERV-W Env Glycoprotein in Human Trophoblast Cell Fusion and Differentiation Direct Involvement of HERV-W Env Glycoprotein in Human Trophoblast Cell Fusion and Differentiation. Society 23, 3566–3574. https://doi.org/10.1128/MCB.23.10.3566
Fuchs, N. V., Kraft, M., Tondera, C., Hanschmann, K.-M., Lower, J., Lower, R., 2011. Expression of the Human Endogenous Retrovirus (HERV) Group HML-2/HERV-K Does Not Depend on Canonical Promoter Elements but Is Regulated by Transcription Factors Sp1 and Sp3. J. Virol. 85, 3436–3448. https://doi.org/10.1128/JVI.02539-10
Fuchs, N. V, Loewer, S., Daley, G.Q., Izsvák, Z., Löwer, J., Löwer, R., 2013. Human endogenous retrovirus K (HML-2) RNA and protein expression is a marker for human embryonic and induced pluripotent stem cells. Retrovirology 10, 115. https://doi.org/10.1186/1742-4690-10-115
Gaedicke, S., Firat-Geier, E., Constantiniu, O., Lucchiari-Hartz, M., Freudenberg, M., Galanos, C., Niedermann, G., 2002. Antitumor Effect of the Human Immunodeficiency Virus Protease Inhibitor Ritonavir: Induction of Tumor-Cell Apoptosis Associated with Perturbation of Proteasomal Proteolysis. CANCER Res. 62, 6901–6908.
Garrison, K.E., Jones, R.B., Meiklejohn, D.A., Anwar, N., Ndhlovu, L.C., Chapman, J.M., Erickson, A.L., Agrawal, A., Spotts, G., Hecht, F.M., Rakoff-Nahoum, S., Lenz, J., Ostrowski, M.A., Nixon, D.F., 2007. T cell responses to human endogenous retroviruses in HIV-1 infection. PLoS Pathog. 3, 1617–1627. https://doi.org/10.1371/journal.ppat.0030165
Gatti, G., Di Biagio, A., Casazza, R., De Pascalis, C., Bassetti, M., Cruciani, M., Vella, S., Bassetti, D., 1999. The relationship between ritonavir plasma levels and side-effects: implications for therapeutic drug monitoring. AIDS 13, 2083–9.

George, M., Schwecke, T., Beimforde, N., Hohn, O., Chudak, C., Zimmermann, A., Kurth, R., Naumann, D., Bannert, N., 2011. Identification of the protease cleavage sites in a reconstituted Gag polyprotein of an HERV-K(HML-2) element. Retrovirology 8, 30. https://doi.org/10.1186/1742-4690-8-30
Gifford, R.J., Blomberg, J., Coffin, J.M., Fan, H., Heidmann, T., Mayer, J., Stoye, J., Tristem, M., Johnson, W.E., 2018. Nomenclature for endogenous retrovirus (ERV) loci. Retrovirology 15, 1–11. https://doi.org/10.1186/s12977-018-0442-1
Goering, W., Ribarska, T., Schulz, W.A., 2011. Selective changes of retroelement expression in human prostate cancer. Carcinogenesis 32, 1484–1492. https://doi.org/10.1093/carcin/bgr181
Göke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., Szczerbinska, I., 2015. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell 16, 135–141. https://doi.org/10.1016/j.stem.2015.01.005
Gonzalez-Hernandez, M.J., Cavalcoli, J.D., Sartor, M.A., Contreras-Galindo, R., Meng, F., Dai, M., Dube, D., Saha, A.K., Gitlin, S.D., Omenn, G.S., Kaplan, M.H., Markovitz, D.M., 2014. Regulation of the Human Endogenous Retrovirus K (HML-2) Transcriptome by the HIV-1 Tat Protein. J. Virol. 88, 8924–8935. https://doi.org/10.1128/JVI.00556-14
Gonzalez-Hernandez, M.J., Swanson, M.D., Contreras-Galindo, R., Cookinham, S., King, S.R., Noel, R.J., Kaplan, M.H., Markovitz, D.M., 2012. Expression of Human Endogenous Retrovirus Type K (HML-2) Is Activated by the Tat Protein of HIV-1. J. Virol. 86, 7790–7805. https://doi.org/10.1128/JVI.07215-11
Götzinger, N., Sauter, M., Roemer, K., Mueller-Lantzsch, N., 1996. Regulation of human endogenous retrovirus-K Gag expression in teratocarcinoma cell lines and human tumours. J. Gen. Virol. 77, 2983–2990. https://doi.org/10.1099/0022-1317-77-12-2983
Gough, D.J., Messina, N.L., Clarke, C.J.P., Johnstone, R.W., Levy, D.E., 2012. Constitutive Type I Interferon Modulates Homeostatic Balance through Tonic Signaling. Immunity 36, 166–174. https://doi.org/10.1016/j.immuni.2012.01.011
Grandi, N., Cadeddu, M., Blomberg, J., Tramontano, E., 2016. Contribution of type W human endogenous retroviruses to the human genome: characterization of HERV‐W proviral insertions and processed pseudogenes. Retrovirology. 13:67. https://doi.org/10.1186/s12977-016-0301-x
Gross, L., 1951. "Spontaneous" leukemia developing in C3H mice following inoculation in infancy, with AK-leukemic extracts, or AK-embrvos. Proc. Soc. Exp. Biol. Med. Soc. Exp. Biol. Med. 76, 27–32. https://doi.org/10.3181/00379727-76-18379
Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y., Pera, R.A.R., Wysocka, J., 2015. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature 522, 221–246. https://doi.org/10.1038/nature14308
Gupta, R., Michaud, H.-A., Zeng, X., Debbaneh, M., Arron, S.T., Jones, R.B., Ormsby, C.E., Nixon, D.F., Liao, W., 2014. Diminished humoral responses against and reduced gene expression levels of human endogenous retrovirus-K (HERV-K) in psoriasis. J. Transl. Med. 12, 256. https://doi.org/10.1186/s12967-014-0256-4
Ham, C., Srinivasan, P., Thorstensson, R., Verschoor, E., Fagrouche, Z., Sernicola, L., Ramos, A., Titti, F., Almond, N., Berry, N., 2010. International multicenter study to assess a panel of reference materials for quantification of Simian

immunodeficiency virus RNA in plasma. J. Clin. Microbiol. 48, 2582–2585. https://doi.org/10.1128/JCM.00082-10
Han, K., Konkel, M.K., Xing, J., Wang, H., Lee, J., Meyer, T.J., Huang, C.T., Sandifer, E., Hebert, K., Barnes, E.W., Hubley, R., Miller, W., Smit, A.F.A., Ullmer, B., Batzer, M.A., 2007. Mobile DNA in Old World Monkeys: A Glimpse Through the Rhesus Macaque Genome. Science (80-.). 316, 238–240. https://doi.org/10.1126/science.1139462
Hanecak, R., Mittal, S., Davis, B.R., Fan, H., 1986. Generation of infectious Moloney murine leukemia viruses with deletions in the U3 portion of the long terminal repeat. Mol. Cell. Biol. 6, 4634–40. https://doi.org/10.1128/MCB.6.12.4634
Hanemann, C.O., 2008. Magic but treatable? Tumours due to loss of Merlin. Brain 131, 606–615. https://doi.org/10.1093/brain/awm249
Hanemann, C.O., Evans, D.G., 2006. News on the genetics, epidemiology, medical care and translational research of Schwannomas. J. Neurol. 253, 1533–1541. https://doi.org/10.1007/s00415-006-0347-0
Hanemann, C.O., Rosenbaum, C., Kupfer, S., Wosch, S., Stoegbauer, F., Müller, H.W., 1998. Improved culture methods to expand schwann cells with altered growth behaviour from CMT1a patients. Glia 23, 89–98. https://doi.org/10.1002/(SICI)1098-1136(199806)23:2<89::AID-GLIA1>3.0.CO;2-Z
Hanke, K., Hohn, O., Bannert, N., 2016. HERV-K(HML-2), a seemingly silent subtenant - but still waters run deep. Apmis 124, 67–87. https://doi.org/10.1111/apm.12475
Hanke, K., Kramer, P., Seeher, S., Beimforde, N., Kurth, R., Bannert, N., 2009. Reconstitution of the Ancestral Glycoprotein of Human Endogenous Retrovirus K and Modulation of Its Functional Activity by Truncation of the Cytoplasmic Domain. J. Virol. 83, 12790–12800. https://doi.org/10.1128/JVI.01368-09
He, T.-C., Zhou, S., da Costa, L.T., Yu, J., Kinzler, K.W., Vogelstein, B., 1998. A simplified system for generating recombinant adenoviruses. Proc. Natl. Acad. Sci. 95, 2509–2514. https://doi.org/10.1073/pnas.95.5.2509
Henzy, J.E., Coffin, J.M., 2013. Betaretroviral envelope subunits are noncovalently associated and restricted to the mammalian class. J. Virol. 87, 1937–46. https://doi.org/10.1128/JVI.01442-12
Hervé, C.A., Lugli, E.B., Brand, A., Griffiths, D.J., Venables, P.J.W., 2002. Autoantibodies to human endogenous retrovirus-K are frequently detected in health and disease and react with multiple epitopes. Clin. Exp. Immunol. 128, 75–82. https://doi.org/10.1046/j.1365-2249.2002.01735.x
Hidesaburo Hanafusa, Teruko Hanafusa, and H.R., 1963. THE DEFECTIVENESS OF ROUS SARCOMA VIRUS. PNAS 12, 572–580.
Hilton, D.A., Hanemann, C.O., 2014. Schwannomas and their pathogenesis. Brain Pathol. 24, 205–220. https://doi.org/10.1111/bpa.12125
Hohn, O., Hanke, K., Bannert, N., 2013. HERV-K(HML-2), the Best Preserved Family of HERVs: Endogenization, Expression, and Implications in Health and Disease. Front. Oncol. 3, 1–12. https://doi.org/10.3389/fonc.2013.00246
Hung, T., Pratt, G.A., Sundararaman, B., Townsend, M.J., Chaivorapol, C., Bhangale, T., Graham, R.R., Ortmann, W., Criswell, L.A., Yeo, G.W., Behrens, T.W., 2015. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. Science (80-.). 350, 455–459. https://doi.org/10.1126/science.aac7442
Hurst, T.P., Magiorkinis, G., 2017. Epigenetic control of human endogenous

retrovirus expression: Focus on regulation of long-terminal repeats (LTRs). Viruses 9, 1–13. https://doi.org/10.3390/v9060130
Hurst, T.P., Magiorkinis, G., 2015. Activation of the innate immune response by endogenous retroviruses. J. Gen. Virol. 96, 1207–1218. https://doi.org/10.1099/jgv.0.000017
Ishida, T., Obata, Y., Ohara, N., Matsushita, H., Sato, S., Uenaka, A., Saika, T., Miyamura, T., Chayama, K., Nakamura, Y., Wada, H., Yamashita, T., Morishima, T., Old, L.J., Nakayama, E., 2008. Identification of the HERV-K gag antigen in prostate cancer by SEREX using autologous patient serum and its immunogenicity. Cancer Immun. 8, 1–10. https://doi.org/081042 [pii]
Jackson, A., Hill, A., Puls, R., Else, L., Amin, J., Back, D., Lin, E., Khoo, S., Emery, S., Morley, R., Gazzard, B., Boffito, M., 2011. Pharmacokinetics of plasma lopinavir/ritonavir following the administration of 400/100 mg, 200/150 mg and 200/50 mg twice daily in HIV-negative volunteers. J. Antimicrob. Chemother. 66, 635–640. https://doi.org/10.1093/jac/dkq468
Jaratlerdsiri, W., Rodríguez-Zárate, C. J., Isberg, S. R., Damayanti, C. S., Miles, L. G., Chansue, N., Moran, C., Melville, L., Gongora, J., 2009. Distribution of endogenous retroviruses in crocodilians. J. Virol. 83, 10305-8. https://doi.org/10.1128/JVI.00668-09
Jern, P., Coffin, J.M., 2008. Effects of Retroviruses on Host Genome Function. Annu. Rev. Genet. 42, 709–732. https://doi.org/10.1146/annurev.genet.42.110807.091501
Jern, P., Sperber, G.O., Blomberg, J., 2005. Use of Endogenous Retroviral Sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. Retrovirology. 2,50.https://doi.org/10.1186/1742-4690-2-50
Jiang, J., Yu, J., Li, J., Li, P., Fan, Z., Niu, L., Deng, J., Yue, B., Li, J., 2016. Mitochondrial genome and nuclear markers provide new insight into the evolutionary history of macaques. PLoS One 11, e0154665. https://doi.org/10.1371/journal.pone.0154665
Jones, R.B., Garrison, K.E., Mujib, S., Mihajlovic, V., Aidarus, N., Hunter, D. V, Martin, E., John, V.M., Zhan, W., Faruk, N.F., Gyenes, G., Sheppard, N.C., Priumboom-brees, I.M., Goodwin, D. a, Chen, L., Rieger, M., Muscat-king, S., Loudon, P.T., Stanley, C., Holditch, S.J., Wong, J.C., Clayton, K., Duan, E., Song, H., Xu, Y., Sengupta, D., Tandon, R., Sacha, J.B., Brockman, M. a, Benko, E., Kovacs, C., Nixon, D.F., Ostrowski, M. a, 2012. HERV-K – specific T cells eliminate diverse HIV-1 / 2 and SIV primary isolates. J. Clin. Invest. 122, 4473–4489. https://doi.org/10.1172/JCI64560.diversity
Jong, S.-M.J., Zong, C.S., Dorai, T., Wang, L.-H., 1992. Transforming Properties and Substrate Specificities of the Protein Tyrosine Kinase Oncogenes ros and src and Their Recombinants. J. Virol. 66, 4909–4918.
Juan, W.C., Hong, W., 2016. Targeting the Hippo signaling pathway for tissue regeneration and cancer therapy. Genes (Basel). https://doi.org/10.3390/genes7090055
Kämmerer, U., Germeyer, A., Stengel, S., Kapp, M., Denner, J., 2011. Human endogenous retrovirus K (HERV-K) is expressed in villous and extravillous cytotrophoblast cells of the human placenta. J. Reprod. Immunol. 91, 1–8. https://doi.org/10.1016/j.jri.2011.06.102
Karamitros, T., Paraskevis, D., Hatzakis, A., Psichogiou, M., Elefsiniotis, I., Hurst, T., Geretti, A.M., Beloukas, A., Frater, J., Klenerman, P., Katzourakis, A., Magiorkinis, G., 2016. A contaminant-free assessment of Endogenous Retroviral

RNA in human plasma. Sci. Rep. 6, 1–12. https://doi.org/10.1038/srep33598
Kato, S., Matsuo, K., Nishimura, N., Takahashi, N., Takano, T., 1987. The entire nucleotide sequence of baboon endogenous virus DNA: A chimeric genome structure of murine type C and simian type D retroviruses. Japanese J. Genet. 62, 127–137. https://doi.org/10.1266/jjg.62.127
Katoh, I., Mírová, A., Kurata, S., Murakami, Y., Horikawa, K., Nakakuki, N., Sakai, T., Hashimoto, K., Maruyama, A., Yonaga, T., Fukunishi, N., Moriishi, K., Hirai, H., 2011. Activation of the Long Terminal Repeat of Human Endogenous Retrovirus K by Melanoma-Specific Transcription Factor MITF-M. Neoplasia 13, 1081-IN42. https://doi.org/10.1593/neo.11794
Katzourakis, A., Magiorkinis, G., Lim, A.G., Gupta, S., Belshaw, R., Gifford, R., 2014. Larger Mammalian Body Size Leads to Lower Retroviral Activity. PLoS Pathog. 10. https://doi.org/10.1371/journal.ppat.1004214
Kent, W.J., 2002. BLAT — The BLAST -Like Alignment Tool. Genome Res. 12, 656–664. https://doi.org/10.1101/gr.229202.
Kessler, A., Wiesner, M., Denner, J., Kämmerer, U., Vince, G., Linsenmann, T., Löhr, M., Ernestus, R.-I., Hagemann, C., 2014. Expression-analysis of the human endogenous retrovirus HERV-K in human astrocytic tumors. BMC Res. Notes 7, 159. https://doi.org/10.1186/1756-0500-7-159
Kim, M., Kim, T., Johnson, R.L., Lim, D.S., 2015. Transcriptional co-repressor function of the hippo pathway transducers YAP and TAZ. Cell Rep. 11, 270–282. https://doi.org/10.1016/j.celrep.2015.03.015
Kim, Y.-J., Lee, J., Han, K., 2012. Transposable Elements: No More "Junk DNA". Genomics Inform. 10, 226–33. https://doi.org/10.5808/GI.2012.10.4.226
Kitamura, Y., Ayukawa, T., Ishikawa, T., Kanda, T., Yoshiike, K., 1996. Human Endogenous Retrovirus K10 Encodes a Functional Integrase. J. Virol. 70, 3302–3306.
Knössl, M., Löwer, R., Löwer, J., 1999. Expression of the human endogenous retrovirus HTDV/HERV-K is enhanced by cellular transcription factor YY1. J. Virol. 73, 1254–61.
Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A.F., Scolnick, E.M., Sigal, I.S., 1988. Active human immunodeficiency virus protease is required for viral infectivity (aspartyl protease/active-site mutation/gag p55/polyprotein processing/in vitro assay). Biochemistry 85, 4686–4690.
Konvalinka, J., Kräusslich, H.-G., Müller, B., 2015. Retroviral proteases and their roles in virion maturation. Virology 479–480, 403–417. https://doi.org/10.1016/J.VIROL.2015.03.021
Kraus, B., Fischer, K., Büchner, S.M., Wels, W.S., Löwer, R., Sliva, K., Schnierle, B.S., 2013. Vaccination Directed against the Human Endogenous Retrovirus-K Envelope Protein Inhibits Tumor Growth in a Murine Model System. PLoS One 8, 1–8. https://doi.org/10.1371/journal.pone.0072756
Kraus, B., Fischer, K., Sliva, K., Schnierle, B.S., 2014. Vaccination directed against the human endogenous retrovirus-K (HERV-K) gag protein slows HERV-K gag expressing cell growth in a murine model system. Virol. J. 11, 58. https://doi.org/10.1186/1743-422X-11-58
Krupovic, M., Blomberg, J., Coffin, J.M., Dasgupta, I., Fan, H., Geering, A.D., Gifford, R., Harrach, B., Hull, R., Johnson, W., Kreuze, J.F., Lindemann, D., Llorens, C., Lockhart, B., Mayer, J., Muller, E., Olszewski, N.E., Pappu, H.R., Pooggin, M.M., Richert-Pöggeler, K.R., Sabanadzovic, S., Sanfaçon, H.,

Schoelz, J.E., Seal, S., Stavolone, L., Stoye, J.P., Teycheney, P., Tristem, M., Koonin, E.V., Kuhn, J.H., 2018. Ortervirales: New Virus Order Unifying Five Families of Reverse-Transcribing Viruses. J. Virol. 92, e00515-18. https://doi.org/10.1128/JVI.00515-18

Kuhelj, R., Rizzo, C.J., Chang, C.H., Jadhav, P.K., Towler, E.M., Korant, B.D., 2001. Inhibition of Human Endogenous Retrovirus-K10 Protease in Cell-free and Cell-based Assays. J. Biol. Chem. 276, 16674–16682. https://doi.org/10.1074/jbc.M008763200

Kuhl, B.D., Cheng, V., Wainberg, M.A., Liang, C., 2011. Tetherin and its viral antagonists. J. Neuroimmune Pharmacol. https://doi.org/10.1007/s11481-010-9256-1

Kumar, S., Bryant, C.S., Chamala, S., Qazi, A., Seward, S., Pal, J., Steffes, C.P., Weaver, D.W., Morris, R., Malone, J.M., Shammas, M.A., Prasad, M., Batchu, R.B., 2009. Ritonavir blocks AKT signaling, activates apoptosis and inhibits migration and invasion in ovarian cancer cells. Mol. Cancer 8, 26. https://doi.org/10.1186/1476-4598-8-26

Küry, P., Nath, A., Créange, A., Dolei, A., Marche, P., Gold, J., Giovannoni, G., Hartung, H.P., Perron, H., 2018. Human Endogenous Retroviruses in Neurological Diseases. Trends Mol. Med. https://doi.org/10.1016/j.molmed.2018.02.007

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M. V, Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N.,

Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E. V, Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J. V, Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., International Human Genome Sequencing, C., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. https://doi.org/10.1038/35057062

Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., Heidmann, T., 2013. Paleovirology of "syncytins", retroviral env genes exapted for a role in placentation. Philos. Trans. R. Soc. B Biol. Sci. 368, 20120507–20120507. https://doi.org/10.1098/rstb.2012.0507

Lavie, L., Kitova, M., Maldener, E., Meese, E., Mayer, J., 2005. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K (HML-2). J. Virol. 79, 876. https://doi.org/10.1128/JVI.79.2.876

Lee, W.J., 2003. Analysis of transcriptional regulatory sequences in the human endogenous retrovirus W long terminal repeat. J. Gen. Virol. 84, 2229–2235. https://doi.org/10.1099/vir.0.19076-0

Lemaitre, C., Harper, F., Pierron, G., Heidmann, T., Dewannieux, M., 2014. The HERV-K Human Endogenous Retrovirus Envelope Protein Antagonizes Tetherin Antiviral Activity. J. Virol. 88, 13626–13637. https://doi.org/10.1128/JVI.02234-14

Lemaître, C., Tsang, J., Bireau, C., Heidmann, T., Dewannieux, M., 2017. A human endogenous retrovirus-derived gene that can contribute to oncogenesis by activating the ERK pathway and inducing migration and invasion. PLoS Pathog. 13, 1–22. https://doi.org/10.1371/journal.ppat.1006451

Li, F., Nellaker, C., Sabunciyan, S., Yolken, R.H., Jones-Brando, L., Johansson, A.-S., Owe-Larsson, B., Karlsson, H., 2014. Transcriptional Derepression of the ERVWE1 Locus following Influenza A Virus Infection. J. Virol. 88, 4328–4337. https://doi.org/10.1128/JVI.03628-13

Li, W., Cooper, J., Zhou, L., Yang, C., Erdjument-Bromage, H., Zagzag, D., Snuderl, M., Ladanyi, M., Hanemann, C.O., Zhou, P., Karajannis, M.A., Giancotti, F.G., 2014. Merlin/NF2 Loss-Driven Tumorigenesis Linked to CRL4DCAF1-Mediated Inhibition of the Hippo Pathway Kinases Lats1 and 2 in the Nucleus. Cancer Cell 26, 48–60. https://doi.org/10.1016/j.ccr.2014.05.001

Li, Y., Golemis, E., Hartley, J.W., Hopkins, N., 1987. Disease specificity of nondefective Friend and Moloney murine leukemia viruses is controlled by a small number of nucleotides. J. Virol. 61, 693–700.

Lindahl, T., Gally, J.A., Edelman, G.M., 1969. Properties of Deoxyribonuclease III from Mammalian Tissues* 244, 5014–5019.

Liu, S.-L., Miller, A., 2007. Oncogenic transformation by the jaagsiekte sheep retrovirus envelope protein. Oncogene 26, 789–801.

https://doi.org/10.1038/sj.onc.1209850
Liu, S.-L., Miller, A.D., 2005. Transformation of Madin-Darby Canine Kidney Epithelial Cells by Sheep Retrovirus Envelope Proteins. J. Virol. 79, 927–933. https://doi.org/10.1128/JVI.79.2.927–933.2005
Lu, X., Sachs, F., Ramsay, L.A., Jacques, P. Étienne, Göke, J., Bourque, G., Ng, H.H., 2014. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat. Struct. Mol. Biol. 21, 423–425. https://doi.org/10.1038/nsmb.2799
Luciw, P.A., Leung, N.J., 1992. Mechanisms of retrovirus replication, in: The Retroviridae, Vol 1. Springer US, Boston, MA, pp. 159–298. https://doi.org/10.1007/978-1-4615-3372-6_5
Ma, H., Ma, Y., Ma, W., Williams, D.K., Galvin, T.A., Khan, A.S., 2011. Chemical Induction of Endogenous Retrovirus Particles from the Vero Cell Line of African Green Monkeys. J. Virol. 85, 6579–6588. https://doi.org/10.1128/JVI.00147-11
Maetzig, T., Galla, M., Baum, C., Schambach, A., 2011. Gammaretroviral vectors: Biology, technology and application. Viruses 3, 677–713. https://doi.org/10.3390/v3060677
Magin-Lachmann, C., Hahn, S., Strobel, H., Held, U., Lower, J., Lower, R., 2001. Rec (Formerly Corf) Function Requires Interaction with a Complex, Folded RNA Structure within Its Responsive Element rather than Binding to a Discrete Specific Binding Site. J. Virol. 75, 10359–10371. https://doi.org/10.1128/JVI.75.21.10359-10371.2001
Magin, C., Löwer, R., Löwer, J., 1999. cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K. J. Virol. 73, 9496–507.
Magiorkinis, G., Blanco-Melo, D., Belshaw, R., 2015. The decline of human endogenous retroviruses: extinction and survival. Retrovirology 12, 8. https://doi.org/10.1186/s12977-015-0136-x
Malassiné, A., Handschuh, K., Tsatsaris, V., Gerbaud, P., Cheynet, V., Oriol, G., Mallet, F., Evain-Brion, D., 2005. Expression of HERV-W Env glycoprotein (syncytin) in the extravillous trophoblast of first trimester human placenta. Placenta 26, 556–562. https://doi.org/10.1016/j.placenta.2004.09.002
Mang, R., Goudsmit, J., van der Kuyl, A.C., 1999. Novel endogenous type C retrovirus in baboons: complete sequence, providing evidence for baboon endogenous virus gag-pol ancestry. J. Virol. 73, 7021–7026.
Manghera, M., Douville, R.N., 2013. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? Retrovirology 10, 16. https://doi.org/10.1186/1742-4690-10-16
Marchi, E., Kanapin, A., Magiorkinis, G., Belshaw, R., 2014. Unfixed Endogenous Retroviral Insertions in the Human Population. J. Virol. 88, 9529–9537. https://doi.org/10.1128/JVI.00919-14
Marzio, G., Tyagi, M., Gutierrez, M.I., Giacca, M., 1998. HIV-1 Tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter. Proc. Natl. Acad. Sci. 95, 13519–13524. https://doi.org/10.1073/pnas.95.23.13519
Mattiuzzo, G., Rose, N.J., Almond, N., Towers, G.J., Berry, N., 2013. Upregulation of TRIM5α gene expression after liveattenuated simian immunodeficiency virus vaccination in Mauritian cynomolgus macaques, but TRIM5α genotype has no impact on virus acquisition or vaccination outcome. J. Gen. Virol. 94, 606–611. https://doi.org/10.1099/vir.0.047795-0

Mayer, J., Ehlhardt, S., Seifert, M., Sauter, M., Müller-Lantzsch, N., Mehraein, Y., Zang, K.D., Meese, E., 2004. Human endogenous retrovirus HERV-K(HML-2) proviruses with Rec protein coding capacity and transcriptional activity. Virology 322, 190–198. https://doi.org/10.1016/j.virol.2004.01.023
Medzhitov, R., 2001. Toll-like receptors and innate immunity. Nat. Rev. Immunol. https://doi.org/10.1038/35100529
Messeguer, X., Escudero, R., Farré, D., Núñez, O., Martínez, J., Albà, M.M., 2002. PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics 18, 333–334. https://doi.org/10.1093/bioinformatics/18.2.333
Mi, S., Lee, X., Li, X. ping, Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., Keith, J.C., McCoy, J.M., 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 403, 785–789. https://doi.org/10.1038/35001608
Michaud, H.-A., de Mulder, M., SenGupta, D., Deeks, S.G., Martin, J.N., Pilcher, C.D., Hecht, F.M., Sacha, J.B., Nixon, D.F., 2014a. Trans-activation, post-transcriptional maturation, and induction of antibodies to HERV-K (HML-2) envelope transmembrane protein in HIV-1 infection. Retrovirology 11, 10. https://doi.org/10.1186/1742-4690-11-10
Michaud, H.-A., SenGupta, D., de Mulder, M., Deeks, S.G., Martin, J.N., Kobie, J.J., Sacha, J.B., Nixon, D.F., 2014b. Cutting Edge: An Antibody Recognizing Ancestral Endogenous Virus Glycoproteins Mediates Antibody-Dependent Cellular Cytotoxicity on HIV-1–Infected Cells. J. Immunol. 193, 1544–1548. https://doi.org/10.4049/jimmunol.1302108
Morandi, E., Tanasescu, R., Tarlinton, R.E., Constantinescu, C.S., Zhang, W., Tench, C., Gran, B., 2017. The association between human endogenous retroviruses and multiple sclerosis: A systematic review and meta-analysis. PLoS One. https://doi.org/10.1371/journal.pone.0172415
Morozov, V.A., Dao Thi, V.L., Denner, J., 2013. The Transmembrane Protein of the Human Endogenous Retrovirus - K (HERV-K) Modulates Cytokine Release and Gene Expression. PLoS One 8. https://doi.org/10.1371/journal.pone.0070399
Moulard, M., Decroly, E., 2000. Maturation of HIV envelope glycoprotein precursors by cellular endoproteases. Biochim. Biophys. Acta - Rev. Biomembr. https://doi.org/10.1016/S0304-4157(00)00014-9
Muir, A., Lever, A.M.L., Moffett, A., 2006. Human endogenous retrovirus-W envelope (syncytin) is expressed in both villous and extravillous trophoblast populations. J. Gen. Virol. 87, 2067–2071. https://doi.org/10.1099/vir.0.81412-0
Nellåker, C., Yao, Y., Jones-Brando, L., Mallet, F., Yolken, R.H., Karlsson, H., 2006. Transactivation of elements in the human endogenous retrovirus W family by viral infection. Retrovirology 3, 44. https://doi.org/10.1186/1742-4690-3-44
Nelson, P., Rylance, P., Roden, D., Trela, M., Tugnet, N., 2014. Viruses as potential pathogenic agents in systemic lupus erythematosus. Lupus 23, 596–605. https://doi.org/10.1177/0961203314531637
Niwa, O., Sugahara, T., 1981. 5-Azacytidine induction of mouse endogenous type C virus and suppression of DNA methylation. Proc. Natl. Acad. Sci. U. S. A. 78, 6290–6294. https://doi.org/10.1073/pnas.78.10.6290
Ogata, T., Okui, N., Sakuma, R., Kobayashi, N., Kitamura, Y., 1999. Integrase of human endogenous retrovirus K-10 supports the replication of replication-incompetent Int- human immunodeficiency virus type 1 mutant. Jpn. J. Infect. Dis. 52, 251–2.

Oldstone, M.B.A., 2014. Molecular Mimicry: Its Evolution from Concept to Mechanism as a Cause of Autoimmune Diseases. Monoclon. Antib. Immunodiagn. Immunother. 33, 158–165. https://doi.org/10.1089/mab.2013.0090

Ono, M., 1986. Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. J. Virol. 58, 937–944.

Ono, M., Yasunaga, T., Miyata, T., Ushikubo, H., 1986. Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J Virol 60, 589–598.

Oricchio, E., Sciamanna, I., Beraldi, R., Tolstonog, G. V., Schumann, G.G., Spadafora, C., 2007. Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. Oncogene 26, 4226–4233. https://doi.org/10.1038/sj.onc.1210214

Osada, N., Hashimoto, K., Kameoka, Y., Hirata, M., Tanuma, R., Uno, Y., Inoue, I., Hida, M., Suzuki, Y., Sugano, S., Terao, K., Kusuda, J., Takahashi, I., 2008. Large-scale analysis of Macaca fascicularis transcripts and inference of genetic divergence between M. fascicularis and M. mulatta. BMC Genomics 9, 90. https://doi.org/10.1186/1471-2164-9-90

Padow, M., Lai, L., Fisher, R.J., Zhou, Y.C., Wu, X., Kappes, J.C., Towler, E.M., 2000. Analysis of Human Immunodeficiency Virus Type 1 Containing HERV-K Protease. AIDS Res. Hum. Retroviruses 16, 1973–1980. https://doi.org/10.1089/088922200750054701

Pan, D., 2010. The hippo signaling pathway in development and cancer. Dev. Cell 19, 491–505. https://doi.org/10.1016/j.devcel.2010.09.011

Parkinson, D.B., Bhaskaran, A., Arthur-Farraj, P., Noon, L.A., Woodhoo, A., Lloyd, A.C., Feltri, M.L., Wrabetz, L., Behrens, A., Mirsky, R., Jessen, K.R., 2008. c-Jun is a negative regulator of myelination. J. Cell Biol. 181, 625–637. https://doi.org/10.1083/jcb.200803013

Payne, L.N., Chubb, R.C., 1968. Studies on the nature and genetic control of an antigen in normal chick embryos which reacts in the COFAL test. J. Gen. Virol. 3, 379–391. https://doi.org/10.1099/0022-1317-3-3-379

Pérot, P., Mugnier, N., Montgiraud, C., Gimenez, J., Jaillard, M., Bonnaud, B., Mallet, F., 2012. Microarray-based sketches of the HERV transcriptome landscape. PLoS One 7. https://doi.org/10.1371/journal.pone.0040194

Perron, H., Dougier-Reynaud, H.L., Lomparski, C., Popa, I., Firouzi, R., Bertrand, J.B., Marusic, S., Portoukalian, J., Jouvin-Marche, E., Villiers, C.L., Touraine, J.L., Marche, P.N., 2013. Human endogenous retrovirus protein activates innate immunity and promotes Experimental Allergic Encephalomyelitis in mice. PLoS One 8, e80128. https://doi.org/10.1371/journal.pone.0080128

Perron, H., Jouvin-Marche, E., Michel, M., Ounanian-Paraz, A., Camelo, S., Dumon, A., Jolivet-Reynaud, C., Marcel, F., Souillet, Y., Borel, E., Gebuhrer, L., Santoro, L., Marcel, S., Seigneurin, J.M., Marche, P.N., Lafon, M., 2001. Multiple sclerosis retrovirus particles and recombinant envelope trigger an abnormal immune response in vitro, by inducing polyclonal Vβ16 T-lymphocyte activation. Virology 287, 321–332. https://doi.org/10.1006/viro.2001.1045

Polavarapu, N., Bowen, N.J., McDonald, J.F., 2006. Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. Genome Biol. 7, R51. https://doi.org/10.1186/gb-2006-7-6-r51

Porter, A.G., Jänicke, R.U., 1999. Emerging roles of caspase-3 in apoptosis. Cell

Death Differ. 6, 99–104. https://doi.org/10.1038/sj.cdd.4400476
Richard A. Gibbs, et al., 2007. Evolutionary and Biomedical Insights from the Rhesus Macaque Genome. Science. 316, 222–234. https://doi.org/10.1126/science.1139247
Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S., Khvorova, A., 2004. Rational siRNA design for RNA interference. Nat. Biotechnol. 22, 326–330. https://doi.org/10.1038/nbt936
Rolland, A., Jouvin-Marche, E., Viret, C., Faure, M., Perron, H., Marche, P.N., 2006. The Envelope Protein of a Human Endogenous Retrovirus-W Family Activates Innate Immunity through CD14/TLR4 and Promotes Th1-Like Responses. J. Immunol. 176, 7636–7644. https://doi.org/10.4049/jimmunol.176.12.7636
Rosenbaum, C., Kluwe, L., Mautner, V., Friedrich, R., M"uller, H., Hanemann, C., 1998. Isolation and characterization of Schwann cells from neurofibromatosis type 2 patients. Neurobiol. Dis. 5, 55–64. https://doi.org/10.1006/nbdi.1998.0179
Rouleau, G.A., Merel, P., Lutchman, M., Sanson, M., Zucman, J., Marineau, C., Hoang-Xuan, K., Demczuk, S., Desmaze, C., Plougastel, B., Pulst, S.M., Lenoir, G., Bijlsma, E., Fashold, R., Dumanski, J., Jong, P. de, Parry, D., Eldrige, R., Aurias, A., Delattre, O., Thomas, G., 1993. Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. Nature 363, 515–521. https://doi.org/10.1038/363515a0
Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D.B., Trono, D., 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature 463, 237–240. https://doi.org/10.1038/nature08674
Rowe, H.M., Trono, D., 2011. Dynamic control of endogenous retroviruses during development. Virology 411, 273–287. https://doi.org/10.1016/j.virol.2010.12.007
Ruggieri, A., Maldener, E., Sauter, M., Mueller-Lantzsch, N., Meese, E., Fackler, O.T., Mayer, J., 2009. Human endogenous retrovirus HERV-K(HML-2) encodes a stable signal peptide with biological properties distinct from Rec. Retrovirology 6, 17. https://doi.org/10.1186/1742-4690-6-17
Santoni, F.A., Guerra, J., Luban, J., 2012. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. Retrovirology 9, 111. https://doi.org/10.1186/1742-4690-9-111
Saresella, M., Rolland, A., Marventano, I., Cavarretta, R., Caputo, D., Marche, P., Perron, H., Clerici, M., 2009. Multiple sclerosis-associated retroviral agent (MSRV)-stimulated cytokine production in patients with relapsing-remitting multiple sclerosis. Mult. Scler. 15, 443–447. https://doi.org/10.1177/1352458508100840
Sauter, M., Schommer, S., Kremmer, E., Remberger, K., Lemm, I.N.A., Buck, M., Best, B., Do, G., 1995. Human Endogenous Retrovirus K10 : Expression of Gag Protein and Detection of Antibodies in Patients with Seminomas 69, 414–421.
Schiavetti, F., Thonnard, J., Colau, D., Boon, T., Coulie, P.G., 2002. A Human Endogenous Retroviral Sequence Encoding an Antigen Recognized on Melanoma by Cytolytic T Lymphocytes A Human Endogenous Retroviral Sequence Encoding an Antigen Recognized on Melanoma by Cytolytic T Lymphocytes 1 5510–5516.
Schmitt, K., Heyne, K., Roemer, K., Meese, E., Mayer, J., 2015. HERV-K(HML-2) rec and np9 transcripts not restricted to disease but present in many normal human tissues. Mob. DNA 6, 4. https://doi.org/10.1186/s13100-015-0035-7

Schulze, K.M., Hanemann, C.O., Muller, H.W., Hanenberg, H., 2002. Transduction of wild-type merlin into human schwannoma cells decreases schwannoma cell growth and induces apoptosis. Hum Mol Genet 11, 69–76.
Speck, N. a, Baltimore, D., 1987. Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. Mol. Cell. Biol. 7, 1101–1110. https://doi.org/10.1128/MCB.7.3.1101.Updated
St Laurent, G., Shtokalo, D., Dong, B., Tackett, M.R., Fan, X., Lazorthes, S., Nicolas, E., Sang, N., Triche, T.J., McCaffrey, T.A., Xiao, W., Kapranov, P., 2013. VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. Genome Biol. 14, R73. https://doi.org/10.1186/gb-2013-14-7-r73
Stauffer, Y., Marguerat, S., Ucla, C., Sutkowski, N., Huber, B., Pelet, T., Conrad, B., 2001. A Model Linking Environment and Autoimmunity. Immunity 15, 591–601.
Stetson, D.B., Ko, J.S., Heidmann, T., Medzhitov, R., 2008. Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity. Cell 134, 587–598. https://doi.org/10.1016/j.cell.2008.06.032
Stocking, C., Kozak, A., 2008. HHS Public Access. Cell Mol Life Sci. 36, 1011–1014. https://doi.org/10.1002/jmri.23741.Proton
Stoye, J.P., 2012. Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nat. Rev. Microbiol. 10, 395–406. https://doi.org/10.1038/nrmicro2783

Stoye, J.P., Coffin, J.M.,1987. The four classes of endogenous murine leukemia virus: structural relationships and potential for recombination. J. Virol. 61, 2659–69.
Subramanian, R.P., Wildschutte, J.H., Russo, C., Coffin, J.M., 2011. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology 8, 90. https://doi.org/10.1186/1742-4690-8-90
Subramanian, S., Kumar, S., 2003. Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. Genome Res. 13, 838–844. https://doi.org/10.1101/gr.1152803
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739. https://doi.org/10.1093/molbev/msr121
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. https://doi.org/10.1093/molbev/mst197
Telesnitsky, A., Goff, S., 1997. Reverse Transcriptase and the Generation of Retroviral DNA, Retroviruses. Cold Spring Harbor Laboratory Press.
Temin, H.M., Mizutani, S., 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature. https://doi.org/10.1101/SQB.1970.035.01.100
Thery, C., Amigorena, S., Raposo, G., Clayton, A., 2006. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. Curr. Protoc. Cell. Biol. 3.22.1-3.22.29. https://doi.org/10.1002/0471143030.cb0322s30
Timmermans, A., Bentvelzen, P., Hageman, P.C., Calafat, J., 1969. Activation of a Mammary Tumour Virus in O20 Strain Mice by X-irradiation and Urethane. J. Gen. Virol. 4, 619–621. https://doi.org/10.1099/0022-1317-4-4-619
Ting, C.N., Rosenberg, M.P., Snow, C.M., Samuelson, L.C., Meisler, M.H., 1992. Endogenous retroviral sequences are required for tissue-specific expression of a

human salivary amylase gene. Genes Dev. 6, 1457–1465. https://doi.org/10.1101/gad.6.8.1457
Tolosa, J.M., Schjenken, J.E., Clifton, V.L., Vargas, A., Barbeau, B., Lowry, P., Maiti, K., Smith, R., 2012. The endogenous retroviral envelope protein syncytin-1 inhibits LPS/PHA-stimulated cytokine responses in human blood and is sorted into placental exosomes. Placenta 33, 933–941. https://doi.org/10.1016/j.placenta.2012.08.004
Tönjes, R.R., Boller, K., Limbach, C., Lugert, R., Kurth, R., 1997. Characterization of Human Endogenous Retrovirus Type K Virus-like Particles Generated from Recombinant Baculoviruses. Virology 233, 280–291. https://doi.org/10.1006/viro.1997.8614
Toufaily, C., Landry, S., Leib-Mosch, C., Rassart, E., Barbeau, B., 2011. Activation of LTRs from different human endogenous retrovirus (HERV) families by the HTLV-1 tax protein and T-cell activators. Viruses 3, 2146–2159. https://doi.org/10.3390/v3112146
Trofatter, J.A., MacCollin, M.M., Rutter, J.L., Murrell, J.R., Duyao, M.P., Parry, D.M., Eldridge, R., Kley, N., Menon, A.G., Pulaski, K., Haase, V.H., Ambrose, C.M., Munroe, D., Bove, C., Haines, J.L., Martuza, R.L., MacDonald, M.E., Seizinger, B.R., Short, M.P., Buckler, A.J., Gusella, J.F., 1993. A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. Cell 72, 791–800. https://doi.org/10.1016/0092-8674(93)90406-G
Turner, G., Barbulescu, M., Su, M., Jensen-Seaman, M.I., Kidd, K.K., Lenz, J., 2001. Insertional polymorphisms of full-length endogenous retroviruses in humans. Curr. Biol. 11, 1531–5.
Ui-Tei, K., Naito, Y., Nishi, K., Juni, A., Saigo, K., 2008. Thermodynamic stability and Watson-Crick base pairing in the seed duplex are major determinants of the efficiency of the siRNA-based off-target effect. Nucleic Acids Res. 36, 7100–7109. https://doi.org/10.1093/nar/gkn902
Uleri, E., Mei, A., Mameli, G., Poddighe, L., Serra, C., Dolei, A., 2014. HIV Tat acts on endogenous retroviruses of the W family and this occurs via Toll-like receptor 4. Aids 28, 2659–2670. https://doi.org/10.1097/QAD.0000000000000477
Utermark, T., Alekov, A., Lerche, H., Abramowski, V., Giovannini, M., Hanemann, C.O., 2003. Quinidine impairs proliferation of neurofibromatosis type 2-deficient human malignant mesothelioma cells. Cancer 97, 1955–1962. https://doi.org/10.1002/cncr.11275
van der Kuyl, a C., Mang, R., Dekker, J.T., Goudsmit, J., 1997. Complete nucleotide sequence of simian endogenous type D retrovirus with intact genome organization: evidence for ancestry to simian retrovirus and baboon endogenous virus. J. Virol. 71, 3666–3676.
Vargas, A., Moreau, J., Landry, S., LeBellego, F., Toufaily, C., Rassart, É., Lafond, J., Barbeau, B., 2009. Syncytin-2 Plays an Important Role in the Fusion of Human Trophoblast Cells. J. Mol. Biol. 392, 301–318. https://doi.org/10.1016/j.jmb.2009.07.025
Vargas, A., Zhou, S., Eťhier-Chiasson, M., Flipo, D., Lafond, J., Gilbert, C., Barbeau, B., 2014. Syncytin proteins incorporated in placenta exosomes are important for cell uptake and show variation in abundance in serum exosomes from patients with preeclampsia. FASEB J. 28, 3703–3719. https://doi.org/10.1096/fj.13-239053
Varki, A., Cummings, R.D., Esko, J., Freeze, H., Stanley, P., Bertozzi, C., Hart, G.,

Etzler, M., 1999. Essentials of Glycobiology, Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Print. Cold Spring Harbor Laboratory Press. https://doi.org/10.1016/S0962-8924(00)01855-9
Vassilev, A., Kaneko, K.J., Shu, H., Zhao, Y., Depamphilis, M.L., 2001. YAP65 , a Src / Yes-associated protein localized in the cytoplasm TEAD / TEF transcription factors utilize the activation domain of YAP65 , a Src / Yes-associated protein localized in the cytoplasm. Genes Dev. 15, 1229–1241. https://doi.org/10.1101/gad.888601
Vogt, P.K., 1967. A virus released by "nonproducing" Rous sarcoma cells. Proc. Natl. Acad. Sci. U. S. A. 58, 801–8.
Volkman, H.E., Stetson, D.B., 2014. The enemy within: Endogenous retroelements and autoimmune disease. Nat. Immunol. 15, 415–422. https://doi.org/10.1038/ni.2872
Wang-Johanning, F., Frost, A.R., Johanning, G.L., Khazaeli, M.B., LoBuglio, A.F., Shaw, D.R., Strong, T. V, 2001. Expression of human endogenous retrovirus k envelope transcripts in human breast cancer. Clin. Cancer Res. 7, 1553–60.
Wang-Johanning, F., Liu, J., Rycaj, K., Huang, M., Tsai, K., Rosen, D.G., Chen, D.T., Lu, D.W., Barnhart, K.F., Johanning, G.L., 2007. Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. Int. J. Cancer 120, 81–90. https://doi.org/10.1002/ijc.22256
Wang-Johanning, F., Radvanyi, L., Rycaj, K., Plummer, J.B., Yan, P., Sastry, K.J., Piyathilake, C.J., Hunt, K.K., Johanning, G.L., 2008. Human endogenous retrovirus K triggers an antigen-specific immune response in breast cancer patients. Cancer Res. 68, 5869–5877. https://doi.org/10.1158/0008-5472.CAN-07-6838
Wang-Johanning, F., Rycaj, K., Plummer, J.B., Li, M., Yin, B., Frerich, K., Garza, J.G., Shen, J., Lin, K., Yan, P., Glynn, S.A., Dorsey, T.H., Hunt, K.K., Ambs, S., Johanning, G.L., 2012. Immunotherapeutic potential of anti-human endogenous retrovirus-k envelope protein antibodies in targeting breast tumors. J. Natl. Cancer Inst. 104, 189–210. https://doi.org/10.1093/jnci/djr540
Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N. V., Schumann, G.G., Chen, W., Lorincz, M.C., Ivics, Z., Hurst, L.D., Izsvák, Z., 2014. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature 516, 405–409. https://doi.org/10.1038/nature13804
Weiss, R., 1967. Spontaneous virus production from "non-virus producing" Rous sarcoma cells. Virology 32, 719–23.
Weiss, R.A., 2006. The discovery of endogenous retroviruses. Retrovirology 3, 67. https://doi.org/10.1186/1742-4690-3-67
Wierer, M., Mann, M., 2016. Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. Hum. Mol. Genet. https://doi.org/10.1093/hmg/ddw208
Xiao, G.-H., Gallagher, R., Shetler, J., Skele, K., Altomare, D.A., Pestell, R.G., Jhanwar, S., Testa, J.R., 2005. The NF2 Tumor Suppressor Gene Product, Merlin, Inhibits Cell Proliferation and Cell Cycle Progression by Repressing Cyclin D1 Expression. Mol. Cell. Biol. 25, 2384–2394. https://doi.org/10.1128/MCB.25.6.2384–2394.2005
Yang, Y.G., Lindahl, T., Barnes, D.E., 2007. Trex1 Exonuclease Degrades ssDNA to Prevent Chronic Checkpoint Activation and Autoimmune Disease. Cell 131, 873–886. https://doi.org/10.1016/j.cell.2007.10.017

Yi, S., Ellsworth, D.L., Li, W.H., 2002. Slow molecular clocks in old world monkeys, apes, and humans. Mol Biol Evol 19, 2191–2198.
Yoshinaka, Y., Katoh, I., Copeland, T.D., Oroszlan, S., 1985. Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. U. S. A. 82, 1618–16122. https://doi.org/10.1073/pnas.82.6.1618
Young, G.R., Eksmond, U., Salcedo, R., Alexopoulou, L., Stoye, J.P., Kassiotis, G., 2012. Resurrection of endogenous retroviruses in antibody-deficient mice. Nature 491, 774. https://doi.org/10.1038/nature11599
Young, N.L., Bieniasz, P.D., 2007. Reconstitution of an infectious human endogenous retrovirus. PLoS Pathog. 3, 0119–0130. https://doi.org/10.1371/journal.ppat.0030010
Yu, F.X., Guan, K.L., 2013. The Hippo pathway: Regulators and regulations. Genes Dev. https://doi.org/10.1101/gad.210773.112
Zeilfelder, U., Frank, O., Sparacio, S., Schön, U., Bosch, V., Seifarth, W., Leib-Mösch, C., 2007. The potential of retroviral vectors to cotransfer human endogenous retroviruses (HERVs) from human packaging cell lines. Gene 390, 175–179. https://doi.org/10.1016/j.gene.2006.08.019
Zeng, M., Hu, Z., Shi, X., Li, X., Zhan, X., Li, X., Wang, J., Choi, J.H., Wang, K., Purrington, T., Tang, M., Deberardinis, R.J., Moresco, E.M.Y., Pedersen, G., Gerald, M., Hedestam, G.B.K., Chen, Z.J., Beutler, B., 2014. HHS Public Access. Science (80-.). 346, 1486–1492. https://doi.org/10.1126/science.346.6216.1486.MAVS
Zhou, F., Krishnamurthy, J., Wei, Y., Li, M., Hunt, K., Johanning, G.L., Cooper, L.J., Wang-Johanning, F., 2015. Chimeric antigen receptor T cells targeting HERV-K inhibit breast cancer and its metastasis through downregulation of Ras. Oncoimmunology 4. https://doi.org/10.1080/2162402X.2015.1047582
Zhou, F., Li, M., Wei, Y., Lin, K., Lu, Y., Shen, J., Johanning, G.L., Wang-Johanning, F., 2016. Activation of HERV-K Env protein is essential for tumorigenesis and metastasis of breast cancer cells. Oncotarget. https://doi.org/10.18632/oncotarget.11455
Zhou, L., Lyons-Rimmer, J., Ammoun, S., Müller, J., Lasonder, E., Sharma, V., Ercolano, E., Hilton, D., Taiwo, I., Barczyk, M., Hanemann, C.O., 2016. The scaffold protein KSR1, a novel therapeutic target for the treatment of Merlin-deficient tumors. Oncogene 35, 3443–3453. https://doi.org/10.1038/onc.2015.404
Zimin, A. V, Cornish, A.S., Maudhoo, M.D., Gibbs, R.M., Zhang, X., Pandey, S., Meehan, D.T., Wipfler, K., Bosinger, S.E., Johnson, Z.P., Tharp, G.K., Marçais, G., Roberts, M., Ferguson, B., Fox, H.S., Treangen, T., Salzberg, S.L., Yorke, J.A., Norgren, R.B., 2014. A new rhesus macaque assembly and annotation for next-generation sequencing analyses. Biol. Direct 9, 20. https://doi.org/10.1186/1745-6150-9-20

12 SUPPLEMENTARY INFORMATION

12.1 Details of analyzed PcEV loci

- **Chr1:55452680-60247** (+ve sense). This is the only locus that appears to have integrated very recently. The LTRs in this locus are identical, suggesting it integrated at least within the last million years (a single substitution would date it to ~1my using our molecular clock). Consistent with being a relatively recent integration, this locus also appears to be heterozygous: we find an identical match to the reconstructed pre-integration site (extending 100nts either side of the TSD) in an unassembled part of the same genome (chrUn_NW_014907418v1), and the earlier rheMac2 assembly from the same animal contains only an identical match to the same reconstructed pre-integration site. Determining homology at this locus is slightly complicated because the locus is within an older ERV integration (designated HERV-17-int in RepeatMasker). Fortunately, this older ERV locus into which PcEv has integrated has diverged markedly from other members of the same HERV-17-int family such that the second best match is only 91% similar, so we can be confident we are comparing the same integration in different builds and in different animals. Unfortunately, the internal region of this locus is both incomplete and possibly even comes from more than one locus: the *gag* is highly degenerate with many frameshifting indels while the (incomplete) *pro-pol* has a potential full-length ORF. As expected from their estimated age, the other intact loci in the rhesus were all also found in the cynomolgus macaque.
- **Chr1:121693612-70205** (-ve sense). The *gag* and *pro-pol* sequences in rheMac2 are identical. The earlier rhesus build also shares the16 amino acid insertion near the start of *env* (as does the cynomolgus genome) and has the same four mismatches between the two LTRs.
- **Chr1:223875631-67155** (-ve sense). We find a full-length locus in the rhe2Mac2 build, except the latter is +sense. This allowed confirmation of all premature stop codons and frameshifting indels in figure 25 and correction of some sequencing errors, e.g. in *pro-pol* a frameshifting loss of one nucleotide from a run of three G in rheMac8 does not occur in rheMac2.

- **Chr2: 120398819-411598** (-ve sense). The *gag* in rheMac8 has many frameshifting indels, none of which are in rheMac2, and the one stop codon in rheMac2 is not in rheMac8 (this sequence does not appear to be in the cynomolgus genome). The *pro-pol* gene has multiple assembly errors, leading to both builds appearing to have long tandem duplications, but neither of these tandem duplications appear in both builds (and neither do any of the premature stop codons or other frameshifting indels); however, the premature stop codon in the centre of the rheMac2 *pro-pol* is also in the cynomologus genome, so we assume that one is real.
- **Chr3:28091955-100392** (-ve sense). The first are 47nts are missing from *gag* in both rhesus builds but the ORF is intact in cynomolgus. Other indels are found only in one of the two rhesus builds.
- **Chr3:28106593-15637** (+ve sense). Contains the only full-length *pro-pol* ORF recovered, although there is a frameshifing indel near the end of *gag* that would throw *pro-pol* out of frame. Scaffold gaps are not repeated in both builds.
- **Chr5:123585628-123593787** (-ve sense) Gaps are not repeated in both rhesus builds.
- **Chr8:14596826-605415** (+ve sense). An old but complete locus with multiple premature stop codons.
- **Chr9:50301699-10531** (+ve sense). This locus is also discussed at length in the main text. It is only the upstream part of pro-pol in rheMac8 that is a problem: both (a) the region containing the 5' LTR plus the start of gag and (b) env plus the 3' LTR were very similar in the three builds. In addition, the macFas5 build has a break in the scaffold near the end of *pro-pol* (which appears to involve some sequence duplication). In the reconstruction of this locus shown in figure 25, we chose not to infer a second a second premature stop codon and a frame-shifting indel in the homologous region of *pro-pol*, although these are present in both rheMac2 and macFas5. We are unable to explain why these substitutions are not also be in the pro-pol region of rheMac8 that our phylogenetic analysis suggests is homologous. We suggest a much more detailed analysis of ERVs the macaque genome sequences is required to resolve this. The two LTRs of this locus in rheMac8 differ by a total of 4 substitutions. In the earlier build they differ by 3 substitutions, only 2 of which are shared in both builds.

- **ChrX:88958264-866804** (-ve sense). We could not find this old locus in the rheMac2 build, although it is in macFas5.
- **ChrY:10651278-9714** (-ve sense). The rheMac2 build and the cynomolgus genome are both from females so we cannot check the rheMac8 sequences of this locus (although the rheMac8 build is from a female it includes the Y chromosome from another individual).

12.2 Interesting observation comparing CA-PcEV ad CA-SIV levels

We might expect to find lower levels of PcEV transcription than SIV, an exogenous retrovirus, since ERVs expression is commonly limited by methylation as shown by the increase in transcription after treatment with the hypomethylating agent 5-azaCytidine, (Depil et al., 2002; Götzinger et al., 1996; Ma et al., 2011; Niwa and Sugahara, 1981). In contrast, we would not expect the exogenous retrovirus to be methylated. However, in our study although the mean number of PcEV copies per 1000 copies GAPDH is lower than that for SIV (37 vs 92), the difference is not statistically significant (Wilcoxon matched-pairs signed rank test, p-value = 0.16).

12.3 Overview of the macaque genome

Our reanalysis of the data presented in (Magiorkinis et al., 2015) show the three ERV lineage that have been copying within the macaque genome in the last 5 million to have similar levels of integration, and thus we assume current activity (Fig. S2). An important discovery here was that Magiorkinis et al. (2015) had erroneously referred to BaEV (Baboon Endogenous Retrovirus) rather than PcEV in the macaque. A recombination event in the baboon gave rise to BaEV, which was first described from several baboon species (Antoinette C. van der Kuyll; John T. Dekker; Jaap Goudsmit, 1995). This recombination event led to BaEV containing the *gag* and *pro-pol* of PcEV and the *env* of SERV (van der Kuyl et al., 1997). We confirmed that BaEV is not actually in the macaque by BLATing the complete BaEV reference sequence (NC_022517) to the macaque genome. The top matches were to PcEV loci shown in figure 25 but only the first ~5000nts of the 8507nts were matched. BLATing the

unmatched 3' region recovered loci we had identified as SERV. Thus, we conclude that while the baboon has both PcEV and BaEV lineages, only the former is present in the macaque.

In humans, the most recently active ERV lineage is HERVK, or more accurately HERVK(HML2), which appears to have been copying in humans at least until 250,000 years ago (Marchi et al., 2014). However, Magiorkinis et al. (2015) concluded that the sister lineage of HERVK in the macaque ceased copying ~5mya and we found the loci to be much more degraded than either HERVK in humans or PcEV in the macaque with could find only a couple of full-length ORFs.

In Fig. S4 we show the three ERV lineages (plus HERVK) among a dendrogram of all macaque ERV loci that integrated roughly since the platyrrhine/catarrhine split. Their relative youth is shown by the number of short branches near the tip (bottom) of the tree. As mentioned in the Introduction, larger-bodied animals tend to have has fewer ERV integrations within the last 10mya (Katzourakis et al., 2014) and in this study the macaque is intermediate between the mouse and humans. Indeed, in that study while the macaque and humans are on the regression line, the mouse has even more recently integrated loci than predicted for its small size. We can further support this closer resemblance of the macaque to the human genome in this respect by comparing dendrograms of the mouse, macaque and human (Fig. S4). The macaque thus appears to offer a model system closer to the human than is the mouse in this respect.

12.4 Note on rheMac8 genome assembly

Rhesus macaque genome assembly rheMac8(Mmul_8.0.1) is from same (female) animal as rheMac2, namely animal 17573 (Zimin et al., 2014). This papers states that the bulk of the sequencing was from a female but that they used an unrelated male for the BAC-end sequencing and to 'aid in selective finishing'. We assume this individual gave the Y chromosome.

12.5 Supplementary figures and tables

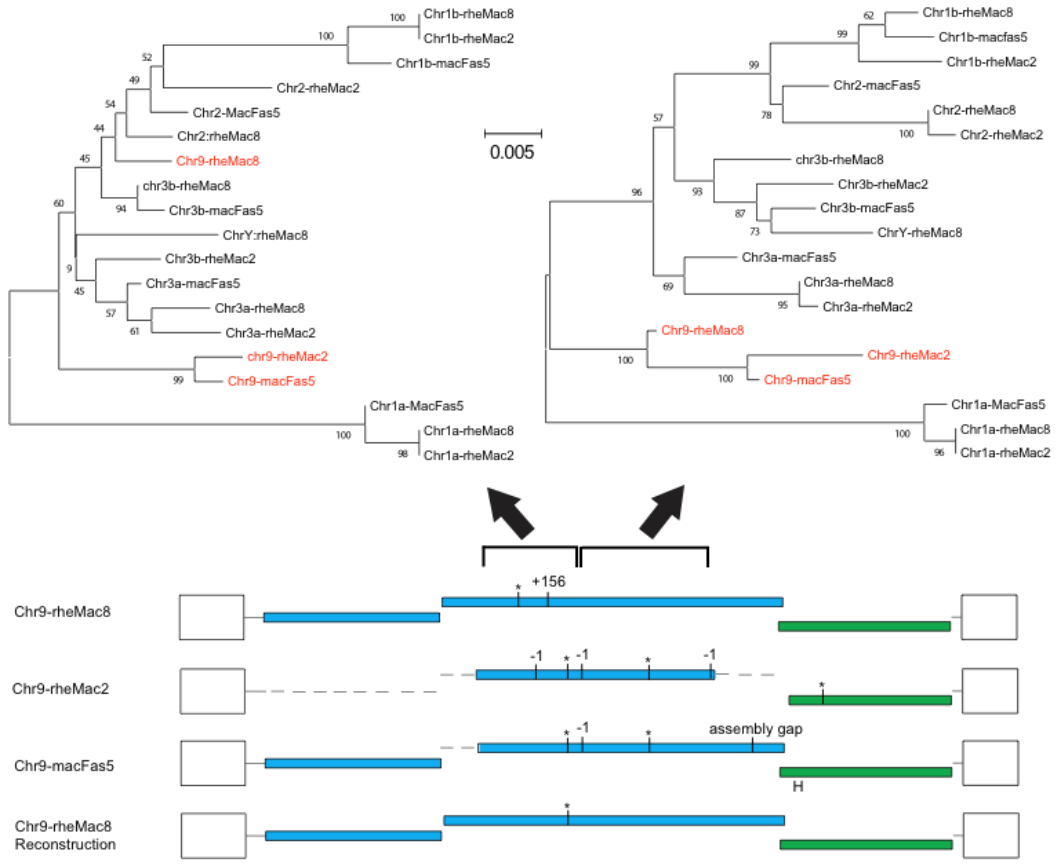


Fig. S1. Reconstruction of the locus chr9:50301699-10531 using homologous sequences in the two rhesus macaque genome assemblies (rheMac2 and rheMac8) and in the cynomolgus macaque genome (macFas5). Phylogenetic trees were built from the 501-1397 and the 1398-2607 regions of the *pro-pol* alignment, with sequences from this locus shown in red.

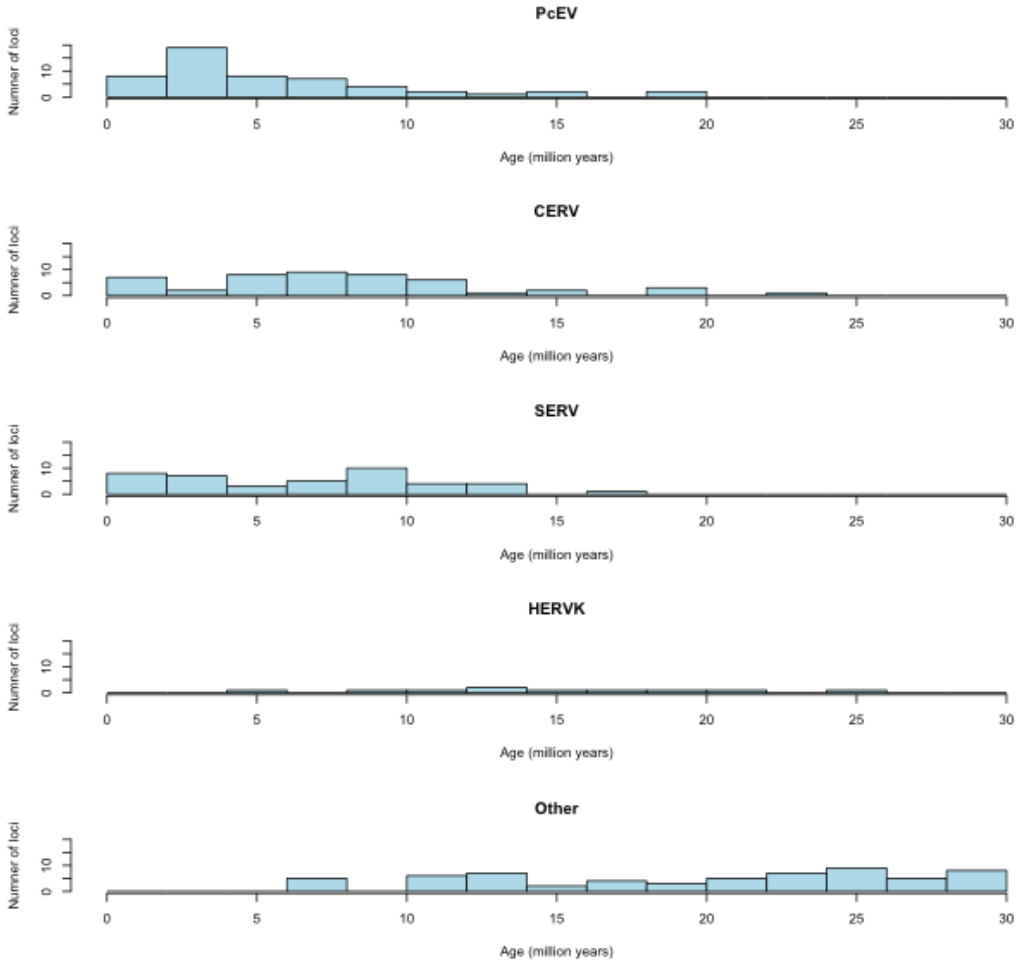


Fig. S2. Histogram showing number of ERV integrations in the rhesus macaque genome (build rheMac2) in 2 million year periods, with integrations dated using LTR divergence. Data re-analysed from Magiorkinis et al. (Magiorkinis et al., 2015).

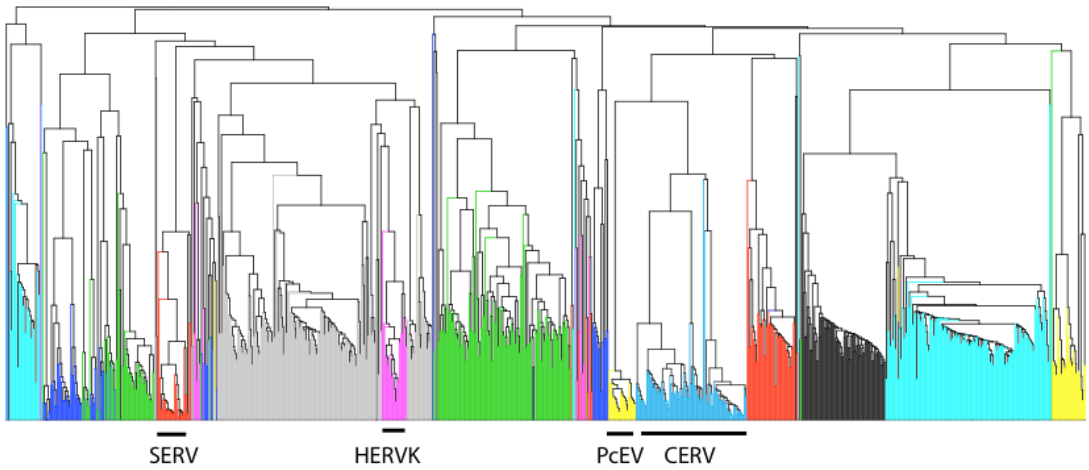


Fig. S3. Dendrogram showing abundance of recently integrated ERV loci in mouse compared to rhesus macaque and human genomes (Magiorkinis et al., 2015). UPGMA dendrogram using pairwise similarity of *pol* gene with very old (>35my) loci excluded. The pair-wise dissimilarity matrix included all having at least 300nt long *pol* match in a region of at least 90% sequence identity with at least one other locus (thus removing loci that would have integrated before the platyrrhine/catarrhine split). The four most recently integrated ERV lineages in the macaque are labelled. We see SERV with short branches in the class II clade (related to alpha and beta exogenous viruses) and a sister group relationship between PcEV and CERV in the class I clade (related to gamma exogenous virsus). CERV shows a long relationship with the macaque genome and probably multiple invasions (there appear to have been multiple invasions of this ERV in the Great Apes.

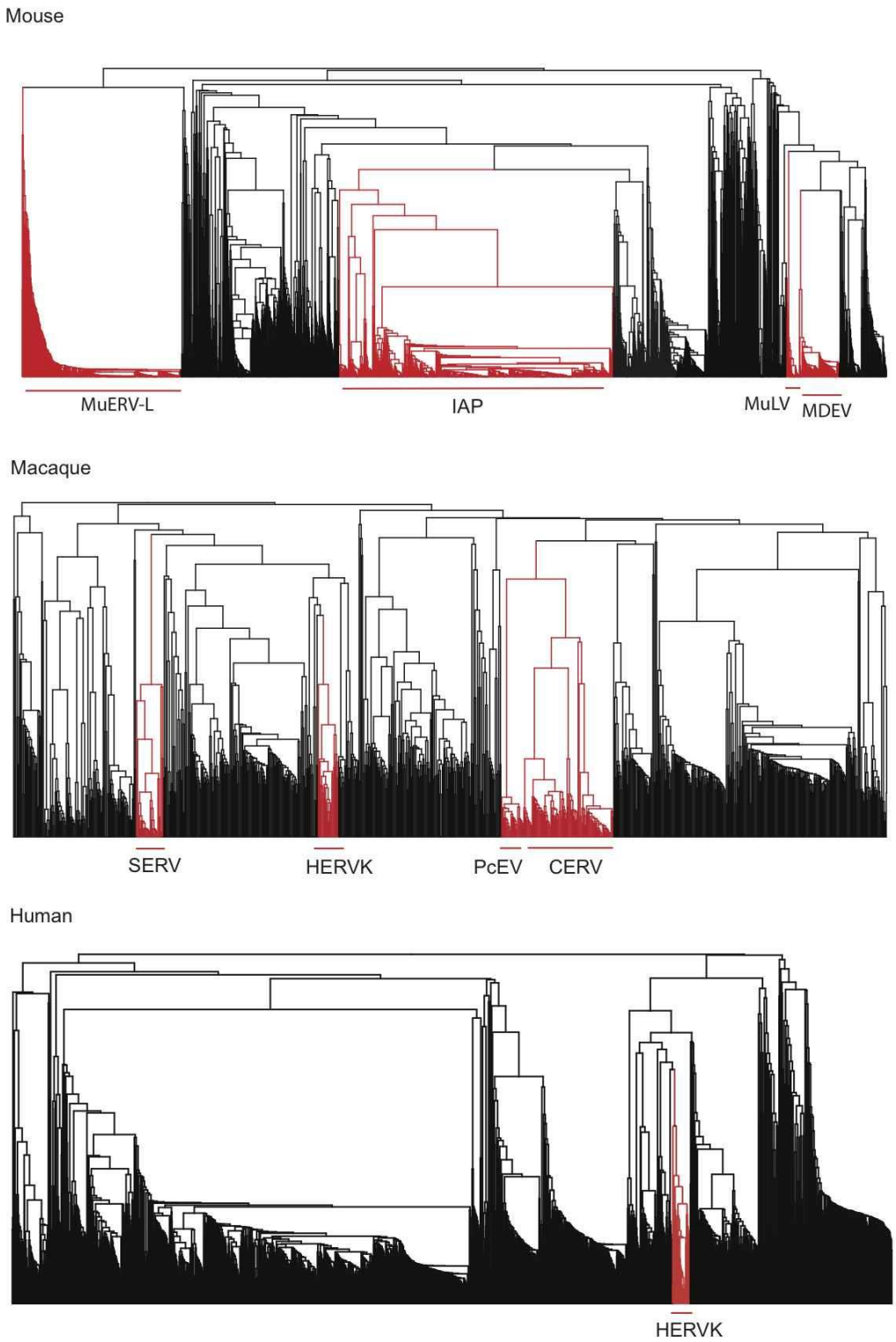
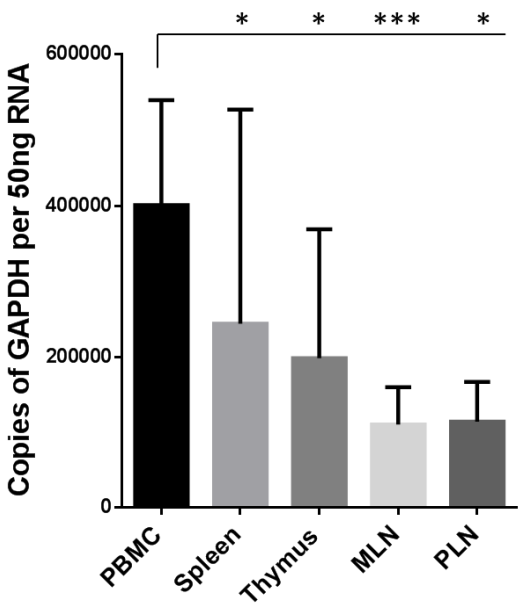


Fig. S4. Dendrogram comparing relative abundance of ERV lineages, shown in red, that contain loci integrating within the last few million years, in the mouse, macaque and human genomes. Details of how the dendrogram were built and their interpretation are given in the legend of fig. S4



GAPDH copies per individual per tissues

	PBMC	spleen	thymus	MLN	PLN
	400000	46600	120000	142000	164000
	472000	182000	43500	229000	139000
	686000	279000	21300	116000	139000
	578000	292000	249000	72700	30000
	452000	294000	246000	89500	99200
	269000	127000	338000	88300	
	396000	318000	652000	115000	
	557000	24800	202000	69400	
	506000	1130000	397000	122000	
	343000	546000	93800	61400	
	300000	121000	78900		
	281000	117000	116000		
	303000	52200	108000		
	185000	64000	109000		
	284000	65200			
Mean	400800	243920	198179	110530	114240
SD	139110	283432	170791	49080	52499
Coef. Var. (%)	34,70808	116,1988	86,18017	44,40423	45,95501

Fig. S5. Comparison of mean copy number of GAPDH among tissues. Kruskal-Wallis followed by Dunn's multiple comparison test revealed that PBMC is the only tissue that display a significantly higher expression of GAPDH compared to spleen, thymus, MLN and PLN that similar expression (non-significantly different). This indicates that inter-tissue comparison is possible between spleen, thymus, MLN and PLN but not PBMC. The table below indicate values per individual per tissues. Mean, standard deviation (SD), coefficient of variation (Coef. Var. (%)) per tissue are provided.

Table S1. PcEV and SIV levels in plasma. Results from RT+ and RT- reactions for each macaque individual are displayed.

		PcEV copies per mL**		SIV copies per ml		
SIV strains	Individuals	RT+	RT-	RT+	RT-	
Naive	E79	No Ct	No Ct	No Ct	No Ct	Extracted from 140uL plasma into 50uL nuclease-free water
	E80	No Ct	No Ct	No Ct	No Ct	
	E81	No Ct	No Ct	No Ct	No Ct	
SIVmac239	E79	No Ct	No Ct	2.60E+06	No Ct	
	E80	No Ct	No Ct	7.24E+05	No Ct	
	E81	No Ct	No Ct	4.23E+06	No Ct	
SIVmacC8	E5	No Ct	No Ct	1.04E+05	No Ct	
	E6	No Ct	No Ct	2.03E+04	No Ct	
	G3	39	No Ct	4.01E+04	No Ct	
	G4	No Ct	No Ct	4.75E+04	No Ct	
	G5	No Ct	No Ct	5.14E+03	No Ct	
	G6	No Ct	No Ct	2.31E+04	No Ct	
SIVmac251	J41	No Ct	No Ct	1.42E+05	No Ct	
	J42	No Ct	No Ct	9.35E+04	No Ct	
	J44	84	No Ct	4.29E+05	No Ct	
	J17	No Ct	No Ct	7.22E+03	No Ct	
	J18	94	No Ct	1.29E+04	No Ct	
SIVsmE660	G25	No Ct	No Ct	2.25E+07	No Ct	
	G26	139	No Ct	3.13E+05	No Ct	
	G27	No Ct	No Ct	2.12E+07	No Ct	
	G28	No Ct	No Ct	3.97E+07	No Ct	
	G7	70	No Ct	9.91E+07	No Ct	
	G8	No Ct	No Ct	2.89E+07	No Ct	
Naive	H17	No Ct	No Ct	No Ct	No Ct	Extracted from 1mL plasma into 100uL nuclease-free water
	H18	No Ct	No Ct	No Ct	No Ct	
SIVmac239	H19	37	No Ct	1,82E+04	No Ct	
	H20	165	No Ct	1,26E+04	No Ct	

**: In a test using plasmas from individuals H17,18,19,20, in the absence of DNAse treatment, we found similar copy numbers between RT+ and RT- reactions: H17(RT+:1.18E+04; RT-:1.27E+04); H18(RT+:7.06E+03; RT-:9.12E+03); H19(RT+:2.39E+04; RT-:2.37E+04); H20(6.96E+03; RT-:8.99E+03). Our observed level of DNA contamination is consistent with our theoretical expectations: assuming the macaque is similar to humans in the level of cell free DNA in the plasma, and given that there are at least 50 intact PcEV copies in each macaque cell, we expected to find at least 5×10^4 (5E+04) DNA copies in the plasma.

pBS-PcEV vector.

Legend
PcEV - ApaI highlighted in Turquoise
Insert in lower case and underlined
T7 promoter *GTAATACGACTCACTATA**G**GGC*
Hind III restriction site **CGATAA**

The reference sequence of the final construct

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGG
CAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGT
GGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTC
GAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAA
GGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCT
TAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATT
ACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGAC
GGCCAGTGAGCGCGC*GTAATACGACTCACTATA**G**GGC*GAATTGGGTACCGGTACCGGGCCccaagccccaatcataattgatctaaa
gcccacggcagtgccgtgtctatcaagcaatatcccatgagccgagaggctcatataggaattcagcagcacattaacaaatttct
agaactcggagtgttgcgaccttgtcgctcgccctggaacactcctcttctgccagtaaaaaagcccggcactcaggattacaggcc
tgtccaagggccCCCCTCGAGGTCGACGTCGACGGTAT**CGATAA**GCTTGATATCGAATTCCTGCAGCCCGGGGGATCCACTAGTTC
TAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGC
TGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAAT
GAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCG
GCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGC
GGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAG
GCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATC
GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATC
TCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAG
GCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAG
TTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGA
TTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAG
GGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATAT
ATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTG
CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCT
CACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCC
AGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCG
TGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCA
AAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGC
ATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC
GGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAAC
GTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAG
CATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAAT
GTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTA
TTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

Table S2. Antibodies used in the thesis

Antibody	Company	Cat#	dilution
Mouse anti-HERV-K (HML2) Env	AMSBio	HERM-1811-5	1:250-1:1000 (WB) 1:50 (IF) 1:100 (IHC)
Mouse anti-HERV-K (HML2) capsid	AMSBio	HERM-1831-5	1:250-1:1000 (WB)
Mouse anti-HERV-K (HML2) Gag	AMSBio	HERM-1841-5	1:250-1:1000 (WB) 1:50 (IF) 1:100 (IHC)
Rabbit anti-phospho ERK	Promega	V803A	1:5000 (WB)
Rabbit anti-phospho AKT	New England Bioloabs	9271	1:500 (WB)
Rabbit anti-phospho FAK^{Y397}	New England Bioloabs	3283	1:500 (WB)
Rabbit anti-ERK	New England Bioloabs	4695	1:500 (WB)
Rabbit anti-AKT	New England Bioloabs	4691	1:500 (WB)
Rabbit anti-FAK	New England Bioloabs	3285	1:500 (WB)
Rabbit anti-Cyclin D1	New England Bioloabs	2922	1:500 (WB)
Rabbit anti-DCAF1/VPRBP	Proteintech	11612-1-AP	1:1000 (WB)
Rabbit anti-CTGF	Abcam	ab6992	1:1000 (WB)
Goat Anti-Mouse IgG (H+L)-HRP Conjugate	Biorad	172-1011	1:10 000-1:20 000 (WB)
Goat Anti-Rabbit IgG (H+L)-HRP Conjugate	Biorad	172-1019	1:10 000-1:20 000 (WB)
Rabbit anti-c-Jun	New England Biolabs	9165	1:100 (IF)
mouse anti-Ki67	Agilent Technologies	M7240	1:100 (IF)
Mouse anti-CD63	Thermofisher	10628D	1:500 (WB)
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermofisher	A11001	1:500 (IF)
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 594	Thermofisher	A11005	1:500 (IF)
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermofisher	A11008	1:500 (IF)
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 568	Thermofisher	A11011	1:500 (IF)