University of Leicester

Epigenetic dynamics of Human Endogenous Retroviruses (HERVs) in human cancer cell lines

Thesis submitted for the degree of

Doctor of Philosophy at the University of Leicester

By

Ibtihal Ismael Al-Shamarti

Department of Genetics and Genome Biology

University of Leicester

January 2018

Epigenetic dynamics of Human Endogenous Retroviruses (HERVs) in human cancer cell lines

Ibtihal Al-Shamarti

Abstract

Transposable elements (TEs) are endogenous components of eukaryotic genomes, constituting 45% of human DNA. The human genome project revealed that human endogenous retroviruses (HERVs) constitute about 8% of the sequence. HERVs are derived from sequences integrated into germ cells during retrovirus infections, up to 25 million years ago. However, most copies of HERVs are defective in multiple ways. The HERV-K family is the youngest family and likely has significant biological activity because of its protein coding capacity.

HERV-K activity may be involved in a variety of cancers, and in particular may play an important role in human melanoma. In this project, HERV activity in melanoma and different cancer cell lines was investigated. Our results showed the HERV-K *pol* gene is expressed in melanoma and in breast, prostate and colon cancers. Furthermore, the response to serum starvation conditions is not simply related to increased expression of HERV-K genes, and changes in cellular phenotype under serum starvation are limited to particular melanoma cell lines, rather than being a general phenomenon.

HERV-K *env* protein expression in melanoma cells was compared to normal primary melanocytes - 1% FBS serum starvation can increase expression of this protein in SKMel5 cells that retain an adherent phenotype. Moreover, *env* protein expression is significantly increased in T47D breast cancer cells under 1% FBS. The analysis of HERV-K LTR methylation state demonstrated that HERV K *env* and *gag* proteins in melanoma cells under 10% FBS and 1% FBS conditions decreased.

The most interesting finding was the detection of 98 candidate loci as novel proviral insertions where previously these loci were annotated as solo LTRs. This result suggests that proviruses are systematically excluded from assemblies and the census of HERV-K proviruses is much greater than represented in assembled genomes. The Genome-wide amplification of proviral sequences combined with Next Generation Sequences (GAPS –NGS) is established as an effective approach to discover new proviral loci.

Acknowledgement

First and foremost, I thank God for giving me patience, perseverance and wisdom during this research and throughout my whole life.

It is my pleasure to thank Dr Richard Badge for organizing a very interesting project. I would also like to thank him for his excellent guidance and his endless kindness and cooperation.

I wish to express my gratitude to Dr Eugene Tulchinsky for his support, valuable ideas, suggestions and interpretation in my project.

I would like to thank Dr Christopher Talbot and Dr Nicola J. Royle for helping me to gain more knowledge and supporting my project.

I would also like to thank Rita Neumann, Dr Nicolas Sylvius and Dr Matthew Blades for helping in iontorrent and RNA seq experiments, and sharing their knowledge.

I would like to thank all the members in lab (Past and present) Maria, Anbhan, Kheloud, Ahmed, Shurooq for helping me in lab whenever it was needed. I would also like to thank Dr Victoria Cotton, Dr Robert Mason, Dr Nicholas Eastley, Daniel Maddison, Nicky Butler and Diana Martin who were always willing to help and useful advice. I would like to thank everyone in lab G18/G19 for their timely help.

Many thanks to members of EMT Lab, especially Hanaa Al-Mahmoodi and Qais Ismaeel, for all their support and excellent company during my time in research. And many thanks to Zamzam, Noura, Ban and Nible for the support, useful advice and lab fun times!

I would also like to thank my close friends Khawla and Russal for their unlimited support during my scientific journey. I would also like to thank Dr Gerald Saldanha for his assistance with the histological aspects of the study.

I would also like to thank Ministry of Higher Education and Scientific Research and University of Kufa to give me the chance to start my PhD studies at the University of Leicester.

I would also like to special and warmest thanks to my family in Iraq my father, mother, brothers, sisters for the continual support that provided the foundation for this work.

Last but not least, my deepest gratitude goes to my husband (Ameen) and kids (Sadeq and Zain) who have been my best friends and they have loved, supported, encouraged, entertained, and helped me get through this agonizing period in the most positive way.

List of abbreviations

APS	Ammonium Persulfate
ATCC	American Type Culture Collection
ATLAS	Amplification typing of L1 active subfamilies
bp, kb	Base pairs, kilo-base pairs
CA	Capsid protein
CEPH	Centre d'Etude du Polymorphisme Humain
DAB	3,3 – diaminobenzidine
DMEM	Dulbecco's Modified Eagle's Medium
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Dexyribonucleic acid
DNMT	DNA methyltransferase
DNMTs	DNA methyltransferases
DSCR	Down syndrome critical region
EDTA	Ethylenediaminetetraacetate
EN	Endonuclease
Env	Envelope
ERVs	Endogenous retroviruses
ERG	ETS-related gene
FBS	Foetal Bovine Serum
FBS	Phosphate Buffered Saline
Fv1	Friend virus susceptibility
g DNA	Genomic DNA
Gag	Group-specific antigen
GATA-1	Erythroid transcription factor
HEMaLP	Primary human epidermal melanocytes, light pigment
HERVs	Human endogenous retroviruses
HIV	Human immunodeficiency virus

HMGS	Human Melanocyte Growth Supplement
HPLC	High performance liquid chromatography
HSV TK	Simplex virus thymidine kinase
HTLV1	Human retrovirusesT cell lymphotropic viruses
ICTV	International Committee on the Taxonomy of Viruses
IF	Immunofluorescence
IHC	Immunohistochemistry
IN	Integrase
JSRV	Jaagsiekte sheep retrovirus
L1	Long Interspersed Elements 1
L1	LINE 1
LINEs	Long interspersed elements
LMP	low melting point
LTR	long terminal repeats
LUMA	luminometric methylation assay
MP	Matrix protein
MelARV	Melanoma associated retrovirus
miRNA	small noncoding RNA
MLV	Murine leukemia virus
MMTV	Mouse Mammary Tumor virus
MPMV	Mason- pfizer monkey virus
MS	Multiple sclerosis
Myr	Million years
NC	Nucleocapsid
NFY	Nuclear transcription factor Y
NGS	Next Generation Sequencing
Np9	Non-structural protein 9
OLB	Oligo Labeling Buffer
ORFs	Open Reading Frames
OSS	Oligo stop solution

OSS	Oligo stop solution
PBS	Primer binding site
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PLZF	Promyelocytic leukemia zinc-finger protein
PNACL	Protein and Nucleic Acid Chemistry Laboratory
Pol	Polymerase
PPT	Polypurine tract
PR	Protease
qPCR	Quantitative polymerase chain reaction
QUMA	QUantification tool for Methylation Analysis software
RA	Rheumatoid arthritis
rec	Accessory protein
RH	Ribonuclease H
RNA	Ribonucleic acid
RT	Reverse Transcriptase
RVLP	Retrovirus-like particles
SDS	Sodium dodecyl sulphate
SETDB1	Histone-lysine N-methyltransferase
SINEs	Short interspersed elements
SLE	Systemic lupus erythematosus
SSC	Sodium chloride citrate
SU	Surface glycoproteins
TAE	Tris-acetate-EDTA
TAS	Ttichostatin A
TBA	Tris-borate-EDTA
TEMED	0.01 ml N,N,N',N'-Tetramethylethylenediamine)
TEs	Transposable elements
TF	Transcription factors

TIRs	Terminal Inverted Repeats
TM	Transmembrane
TPRT	Target site primed reverse transcription
TSD	Target site duplications
UTR	Untranslated region
VLPs	Virus-Like Particle
VNTR	Variable number of tandem repeats
WB	Western Blot

List of contents

List of Figures

List of Table

1 Introduction

1.1 Repetitive content of the human genome

Mobile elements are endogenous components of the genomes of eukaryotes and prokaryotes (Kazazian, 2004). Mobile elements or transposable elements (hereafter TEs) constitute 45% of the human genome DNA (Lander *et al.*, 2001) and are defined as DNA sequences which can copy themselves from one place to a new locus within their host cells (Kazazian, 2004). Although it has been believed that TEs are "Junk DNA", a number of recent studies have demonstrated that transposable elements can play an important role in aspects of human genome evolution and function (Deininger *et al.*, 2003). Moreover, the wide distribution of TE sequences in genomes can provide a useful resource for analyzing genome evolution (Tomilin, 1999). TEs can be divided into two classes as illustrated in figure 1.1, the first class being, retrotransposons or (RNA transposons), which use a "copy and paste" process to replicate. This mechanism involves reverse transcription of an RNA intermediate, and so is called retrotransposition. In contrast, TEs that move by a "cut and paste" mechanism via a DNA intermediate are referred to as DNA transposons (the second major class in Figur1.1A (Munoz-Lopez & Garcia-Perez, 2010).

Another important classification is that between non-autonomous and autonomous elements. This classification depends on whether the TE moves by utilizing enzymes that they encode (autonomous) or parasitize autonomous elements by using their enzymes (non- autonomous) (Sheen *et al.*, 2000; Munoz-Lopez & Garcia-Perez, 2010).

1.1.1 DNA transposons

DNA transposons are found in bacteria, insects, and worms and as fossils in human genomes. In human genomes, they occupy 3% of the DNA (Kazazian, 2004). DNA transposons can be excised from one locus and integrated into other locus by a "cut and paste" mechanism (Chuong *et al.*, 2017).

Kazazian (2004) mentioned that new insertions of DNA transposons generally occur near the original location of the element. This phenomenon is called "local hopping". DNA transposons contain a transposase gene that is usually bounded by two Terminal Inverted Repeats (TIRs). The transposase interacts with these TIRs to cut the transposon's DNA and then catalyzes the "pasting" (insertion) of the DNA into a new locus. At the insertion site the gaps left by the insertion process are filled by host DNA repair enzymes. This situation leads to directly duplicated sequence at the insertion point, known as target site duplications (TSD).

Recombinase enzymes (DNA ligase) are also used to mobilize segments of DNA. These enzymes can be found in two different classes. The first class of recombinases utilizes a molecularly conservative reaction that does not require energy, and the second class are transposases enzymes which require energy to break and reform phosphodiester bonds (Katoh & Kurata, 2013).

1.1.2 Retrotransposons

Retrotransposons are mobilized by reverse transcription of an mRNA intermediate which is then integrated into another locus in the genome. It is important to recognize that there are two classes of retrotransposons: those which are bounded by long terminal repeats (LTR elements) and those which lack LTRs (non-LTRelements) (Munoz-Lopez & Garcia-Perez, 2010; Chuong *et al.*, 2017).

More recent studies have revealed that LTR elements include human endogenous retroviruses (HERV) which constitute about 8% of the human genome, which is the subject of this thesis, and will be considered in detail later (Lander *et al.*, 2001) . Non-LTR retrotransposons are further divided into groups: - (SINEs) Short Interspersed Elements and LINEs Long Interspersed Elements (typified by L1, L2 and L3 elements in humans) (Lander *et al.*, 2001).

1.1.2.1 Non-LTR Retrotransposons

1.1.2.1.1 SINEs

SINES are short sequence approximately 100-300bp in length that constitute of 13% of the human genomes (Lander *et al.*, 2001). These SINEs often use LINEs proteins to mobilize because they are not large enough to encode proteins for themselves, and hence are non-autonomous retrotransposons (Deininger *et al.*, 2003). SINEs generally include an internal RNA polymerase III promoter. Alu elements are the main active SINE elements in human genomes. Alu is derived from the 7SL RNA, via duplication event, and are mobilized by the L1 machinery (Deininger *et al.*, 2003; Dewannieux *et al.*, 2006). Alu has expanded to 10% of the human genome and a new insertion is estimated to occur in 1 in every 30 human births (Landry *et al.*, 2001). Alu elements are approximately 280 bp in length and include two monomers separated by an A-rich tract ending with a poly (A)-Tail. There is also evidence that Alu insertions cause around 1in every 1000 germline mutations, especially those elements from the evolutionarily young Alu Ya5 and Alu Yb8 subfamilies (Deininger *et al.*, 2003).

1.1.2.1.2 LINE

LINEs are autonomous non-LTR retrotransposons found in many genomes. LINE-1 (hereafter L1) elements, the currently active family, comprise 17% of the human genome. Other ancient inactive LINEs also occur in the human genome (LINE2 and LINE3) as molecular fossils, occupying only 4% of human DNA. However, L1 elements are currently actively retrotransposing and so frequently show presence or absence variation between individuals. It should be noted that of the approximately 500,000 L1 copies in the human genome, only 80~100 potentially active elements occur in the average diploid human genome (Brouha *et al.*, 2003). The reason for this, noted by Brouha *et al.* (2003), is that 99.9% of L1s are inactive because of 5′ truncations, insertions and inversions. L1 elements that are inserted into new genomic loci by retrotransposition, even if truncated or mutated may still disrupt genes and produce different types of disease (Kazazian, 2004).

A full length L1 element is approximately 6 kb long containing a 910bp 5'untranslated region (5′ UTR) and a 205 bp 3′ untranslated region (3′ UTR) that ends with poly A tail. Between the UTRs are two open reading frames (ORFs) (Beck *et al.*, 2010) called ORF1 and ORF2, which play a major role in L1 retrotransposition. ORF1 encodes a protein (40 KDa) which acts as an RNA –binding protein required for retrotransposition. On other hand, ORF2 encodes a protein (150 KDa), also necessary for L1 retrotransposition which encodes endonuclease (EN) and reverse transcriptase (RT) functions (Beck *et al.*, 2010).

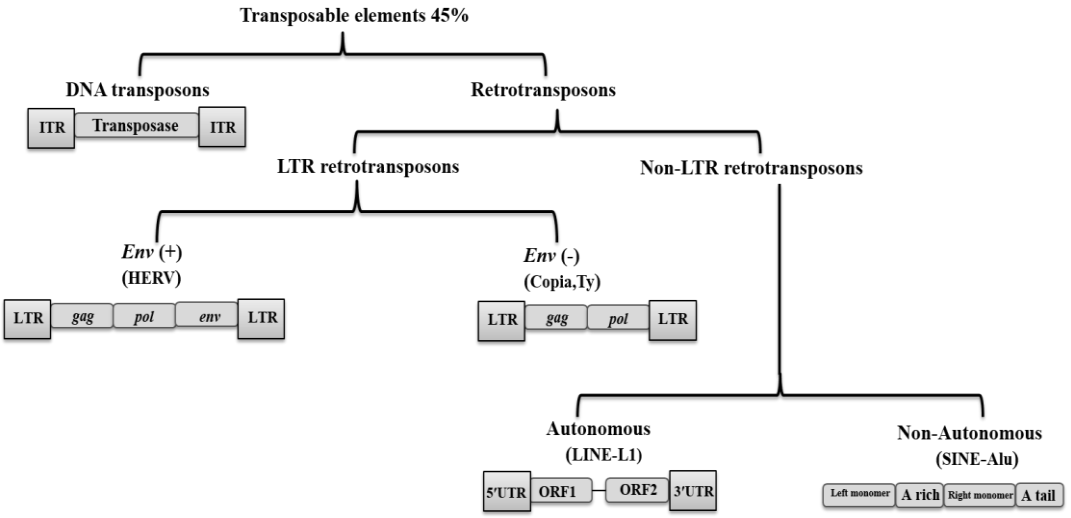


Figure 1.1: Classification of Mobile elements

Mobile elements or TEs constitute about 45% of the human genome (Lander *et al*., 2001). TEs can be divided into two classes: DNA transposons and retrotransposons. Retrotransposons contain two types which are non-LTR retrotransposons, such as SINE, e.g., Alu elements, and LINE, e.g., LINE-1 (L1). LTR retrotransposons, or (endogenous retroviruses (ERVs)) are around 8% of the human genome, for examples HERV, and also LTR retrotransposons divided into types depending on presence or absence of an *env* gene.

1.1.2.2 LTR Retrotransposons (Endogenous retroviruses) ERV

LTR retrotransposons, most often derived from retroviruses, constitute a significant fraction of many genomes. LTR retrotansposon transcripts are converted into DNA by reverse transcription (RT) to form a cDNA which is then inserted into the host DNA genome, adding another copy of the retrotransposon to the genome (Havecker *et al.*, 2004). This process is a type of retrotransposition, but reverse transcription and integration are spatially and temporally separate processes (Figure1.2). It is generally noted that retroviruses contain the same structural and enzymatic portions as LTR retrotransposons.

They include a *gag* genes which encodes a viral particle coat protein, and a *pol* gene that contain reverse transcriptase (RT), ribonuclease H (RH) and integrase (IN) activities, with these gene flanked by long terminal direct repeats (LTR). However, LTR

retrotansposons lack the *env* genes of infectious retroviruses and, as a result have lost the ability to infect new cells (Kassiotis, 2014).

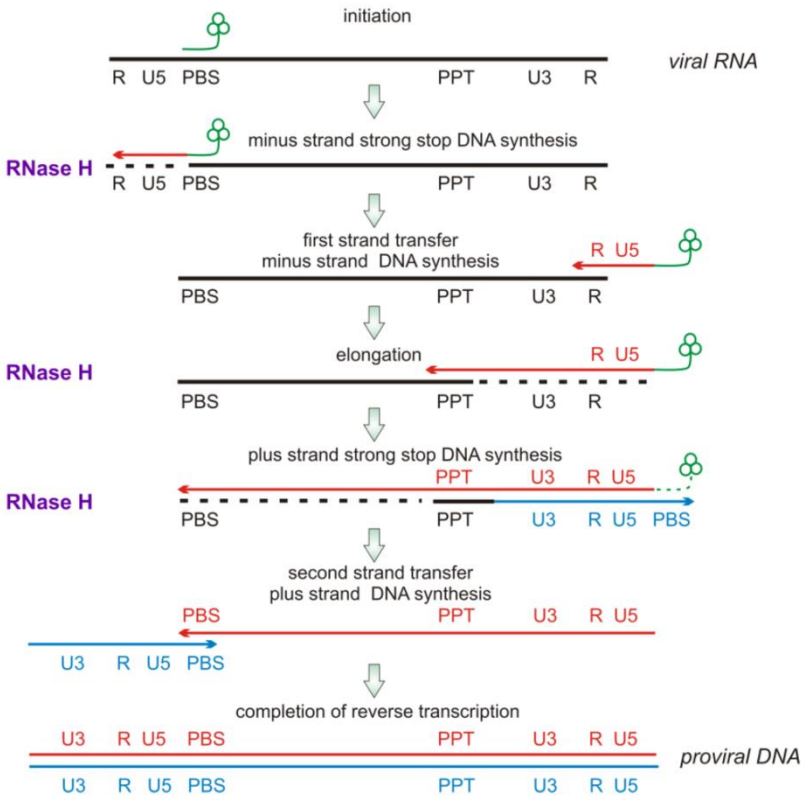


Figure 1.2: Process of cDNA synthesis and formation of LTRs

This process of the provirus RNA genome makes double stranded cDNA. First strand cDNA synthesis (DNA strand in red) is initiated using a cellular tRNA annealed to the Primer binding site (PBS). The RNA strand is degraded by RNase H enzyme. First strand transfer to 3′LTR of the viral genome to extend the newly formed DNA. Transfer is mediated by identical sequence of the R region. Then, RNase H digests all template RNA except Polypurine tract (PPT). The PPT is used as a primer for second strand DNA synthesis. RNase H removes the tRNA and the PPT. PBS in the second strand transfers to the PBS the first strand of first strand cDNA. As a result, both strands are extended to form double-stranded DNA. (From Ilina *et al.*, 2012).

The retroviruses as a class (RV) contain a diverse range of viruses from a variety of host species, which include many types of animals such as birds, mice and pigs (Lisanti *et al.*, 2013). The human genome includes approximately 400,000 loci showing evidence of germ line integration by many types of retroviruses in the primate lineage (Chuong *et al.*, 2017; Cegolon *et al.*, 2013). Approximately 8% of human genome is derived from human endogenous retrovirus (HERV) sequences, which were integrated into germ cells

by exogenous retrovirus infection from around 25 million years ago (Myr) (Cegolon *et al.*, 2013). The integrated viral sequence is called a provirus and is transmitted vertically within the host genome rather than via horizontal transmission by infection (Katoh & Kurata, 2013). It is assumed that the human genome has continued to accumulate HERVs after separation from the Old World monkey lineage, and they have increased in number over primate evolution (Bannert & Kurth, 2006).

1.2 Origin

Generally, ERVs have been found in many different types of vertebrate genomes, (Bannert & Kurth, 2006). Some studies used the mouse to investigate ERVs in the laboratory (Weiss, 2013; Landry *et al.*, 2001; Kalyanaraman *et al.*, 1982). These experiments demonstrated that Mouse Mammary Tumor virus (MMTV) can alternate between vertical and horizontal inheritance, being transmitted by breast-feeding and germ line insertion, respectively. Thus, it can be found as an exogenous or endogenous virus (Kalyanaraman *et al.*, 1982). Also, the murine leukemia virus (MLV) was shown to infect many mouse strains and (Mouse Genome Sequencing Consortium *et al.*, 2002). Kalyanaraman *et al*. (1982) showed that the major enzyme of MLV was a reverse transcriptase. As a result, it became clear that ERVs could potentially cause human diseases such as cancer. In1982 Gallos's group discovered two human retroviruses, T cell lymphotropic viruses (HTLV1) and (HTLV2) (Kalyanaraman *et al.*, 1982; Gallo *et al.*, 1983).

These were the first identified human retroviruses. Late in 1983, HIV was identified and shown to cause AIDS (Gallo *et al*., 1983). Moreover, the first sequences of HERVs were identified in 1987 by using a probe to the polymerase gene in genomic southern

blots, showing these were integrated in the genome. HERV expression has been associated with a variety of cancers, for example melanoma and cancer of the breast, prostate and ovaries (Moles *et al.*, 2005; Muster *et al.*, 2003). Buscher *et al*. (2005), found HERV-K *gag* and *env* proteins were expressed in melanoma cells, while normal melanocytes did not express these proteins. Also, Rastogi *et al*., (2016) found a relatively higher expression level of HERV-K *gag* protein in patients who have prostate cancer, compared to healthy individuals.

1.3 Life cycle of ERVs

The life cycle of an ERV starts with the infection of the host germline by an exogenous retrovirus. As part of the infection cycle the viral RNA genome is converted into cDNA by reverse transcriptase and the cDNA integrated into chromosomal DNA, as a provirus, illustrated in figure 1.3. As these events occur in the germline, the offspring of the infected individual will carry the new ERV provirus as part of their genome. The provirus is subsequently transmitted vertically from parents to offspring in a stable Mendelian fashion by germline transmission (Mastrangelo *et al.*, 2009). After germline transmission occurs the fate of the insertion becomes variable:

It could be lost from the population by genetic drift while other insertions might be fixed within the population and inherited for a long time (Gifford & Tristem, 2003). In general, when exogenous retroviruses are inserted into human DNA, they are called HERVs. HERV can significantly affect the host in two ways. Firstly, HERV insertion within a gene can cause a mutation that disrupts gene expression or function. Secondly, the expression of HERV genes can cause disease (Best *et al.*, 1997).

The reasons for non-infectious transmission of HERV have been attributed to age and activity of virus, and selective pressure (Kassiotis, 2014). On the other hand, several studies have demonstrated HERVs may have retained the capacity for viral infection due to maintenance of the ORFs by positive selection (Belshaw *et al*., 2004)

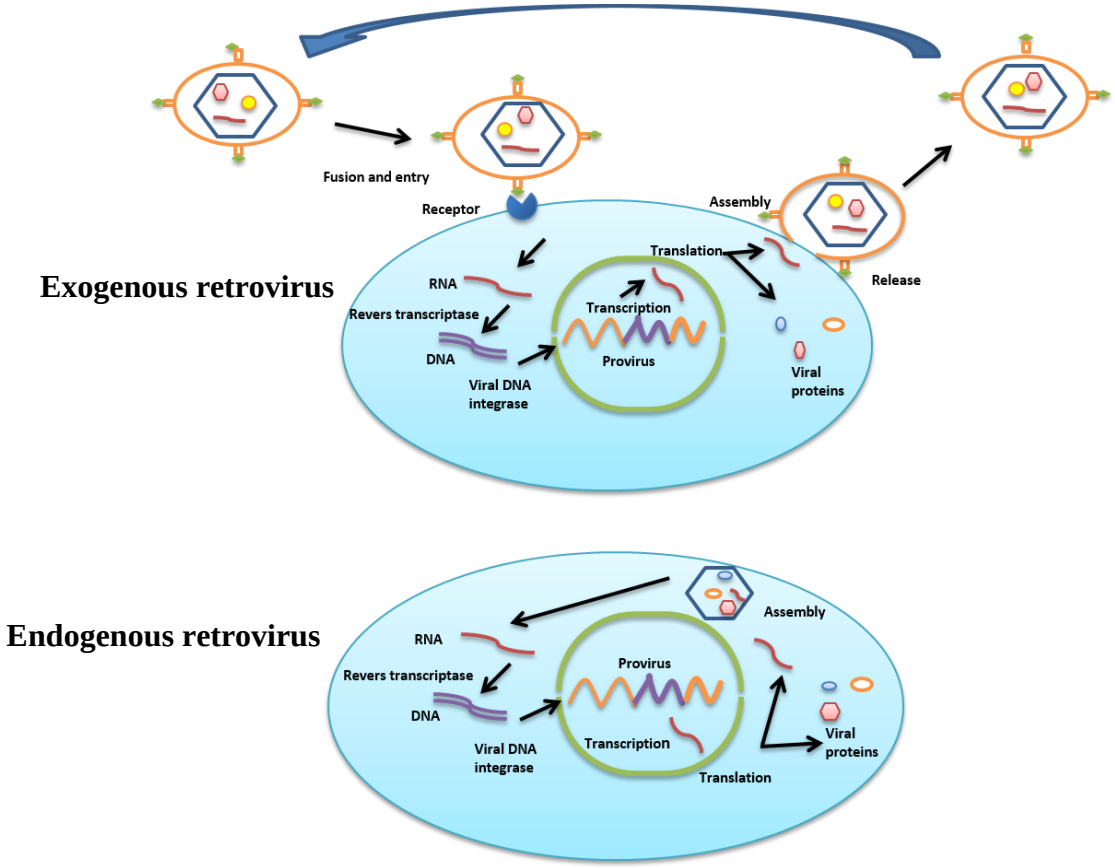


Figure 1.3: The life cycle of retroviruses

(**Upper panel**) Exogenous retroviruses infect cells by a specific receptor on the cell surface. In the cytoplasm, viral RNA is converted into cDNA by reverse transcription (RT), and the cDNA transits through the cytoplasm via nuclear entry to insert into the host DNA genome, to generate a provirus. Transcription of the provirus by RNA polymerase II is followed by translation into viral proteins. Finally, the viral capsid is assembled and released from the cell surface (exogenous retroviruses). (**Lower panel**) In the case of endogenous retroviruses the extracellular phase has been lost, although capsid assembly may occur in the cytoplasm form virus-like-particles (VLPs).

1.4 Structure

Human endogenous retrovirus (HERV) structure constitutes the LTRs, flanking the protein-coding genes *gag*, *pol* and *env* as illustrated in Figure (1.4) (Muster *et al.*, 2003).

The abbreviation *gag* means group-specific antigen, *pol* refers to polymerase while *env* means envelope. The three ORFs are flanked by the 5′ and 3′ LTRs. Not all HERV copies are intact; in the extreme case recombination occurs between the two LTRs to form a single LTR (Solo LTR). This results in removal of the internal segment of the provirus (Figure 1.5) (Kassiotis, 2014; Suntsova *et al.*, 2015).

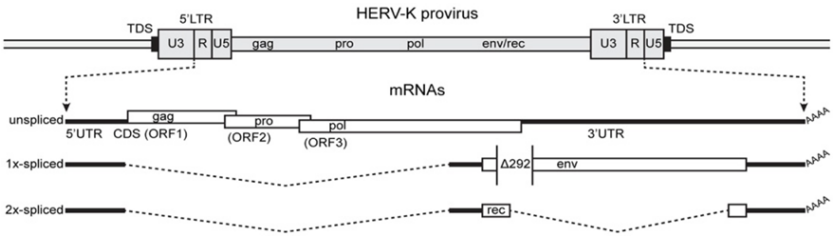


Figure 1.4: Structure of a HERV-K provirus and its spliced mRNAs

The structure of a HERV-K provirus inserted into human genome is shown with target duplicated sequence (TDS) flanking the 5′ and 3′ LTRs. The splicing events resulting in HERV-K type 1(np9) and type 2(rec) specific mRNA. The two types differ by a 292 basepair deletion, from (Agoni *et al.*, 2013).

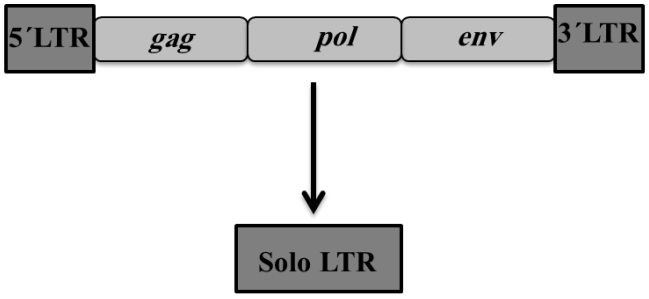


Figure 1.5: Solo-LTRs Schematic structure of HERV-K

Schematic illustration of the structures of solo LTR generated whether from recombination between the two LTR, or deletion of the internal regions of HERV-K structure.

1.4.1 Protein coding ORFs

The *gag* gene encodes multiple proteins, including nucleocapsid proteins (NC), capsid proteins (CA) and matrix proteins (M), while the *env* gene encodes transmembrane(TM)

and surface glycoproteins (SU) (Singh *et al.*, 2009). The *gag* proteins responsible for form virus-like particles in cell, while the *env* proteins responsible for infused into host cell membrane and entry of virus inside cell. In the same time, TM acts as immunosuppressive, thus it could able to contribute in immune system. Moreover, *env* proteins may act as defense against exogenous infection retroviruses by receptor interference signal (Nadeau *et al.*, 2015).

The largest ORF is *pol* because it includes essential activities for viral replication such as reverse transcriptase (RT), RNAse H activity. While, integrase (IN) responsible for integrated viral DNA into sequence of host DNA, and protease (PR) responsible for responsible for splitting the *gag-pol* polyproteins during maturation (Gifford & Tristem, 2003). The primer binding site (PBS) is located at the 5′ end between LTR and *gag*, which is important to start reverse transcription process. Whereas the polypurine tract (PPT), is situated between *env* and 3′ LTR, PPT, is essential to finish transcription process. Some HERVs are able to produce proteins such us *gag* and *env* proteins which can affect gene expression, and which may result in altered gene function (Jern *et al.*, 2005).

It has been noticed that the *gag* and *pol* genes, in some HERVs, have a frame shift between them (Gao *et al.*, 2003). Thus it is clear that ribosomal frame shifting is an important step for HERVs to produce *gag-pol* proteins. The -1 reading frame shift at the *gag-pol* sequence boundary alters the reading frame sequence, such that *gag* and *pol* sequences are expressed as a fusion proteins. However, the frame shift process does not occur in some HERVs, and as a result only the *gag* proteins are expressed. The *env* gene has the potential ability to encode a transmembrane protein, which allows the packaged virus to bind to and infect a new cell (Krishnamurthy *et al.*, 2015).

However, contemporary HERVs have lost these genes, and so can no longer propagate by infecting other cells (de Parseval & Heidmann, 2005).

Some HERVs contain an UTPase gene whose role is proposed to protect a new DNA provirus from toxic mis-incorporation of uracil (Jern *et al.*, 2005). However, other HERVs do not have a UTP gene, and so they are dependent on host enzymes for prevent uracil incorporation into host gDNA (Cegolon *et al.*, 2013). Although, the most of HERVs genes have highly mutation and defective on ORF in genome DNA, HERVs still have intact ORF which produced different proteins and potentially influence on gene expression

1.4.2 LTR region

The LTR is an interesting region of HERVs because it acts as a promoter for high-level expression of the viral sequence (Jern *et al.*, 2005). LTRs represent approximately 300 - 1300 bp of the viral genomes and consist of subdomains organized in a standard order: U3-R-U5, as illustrated in Figure 1.2. The U3 part includes a TATA box and transcriptional regulatory sequences (Katoh & Kurata, 2013). Also, U5 includes 75-200 bases of unique sequence that impacts on LTR activity and viral gene expression (Manghera & Douville, 2013). The R region also contains regulatory factors that induce transcription. The R region is retained in mammalian apparent LTR retrotransposons (MaLRs), an LTR element derivative which occupies approximately 3.6%of the human genome (Kassiotis, 2014). The MaLR structure includes the 5' LTR-ORF-LTR 3' but lacks the *gag-pro-pol* and PBS sequences. It is proposed that these sequences enable MaLRs to be transcribed and parasitize the retrotransposition machinery of proteins encoded by ERVs. As result, MaLRs, are non-autonomous retrotransposons. This

situation, where autonomous LTR retrotransposons support the expansion of non-autonomous deletion derivatives is observed in other animal retroelements, for example Gypsy/Ty3 and Copia/Ty1 (Katoh & Kurata, 2013).

1.5 Classification

According to the International Committee on the Taxonomy of Viruses (ICTV) retroviruses consist of seven families, based on gene sequence data. These families are Alpharetrovirus, Betaretrovirus, Deltaretrovirus, Gammaretrovirus, Epsilonretrovirus, Supmavirus and Lentivirus (Gifford & Tristem, 2003) .

The Betaretrovirus family has been investigated intensively in mammals. The Mason-Pfizer Monkey Virus (MPMV), endogenous and exogenous mouse mammary tumor virus (MMTV) and Jaagsiekte sheep retrovirus (JSRV) are all well studied (Gifford & Tristem, 2003). In addition, it is an important family because Betaretroviruses give rise to Class II HERVs including the HERV-K family, which is the principle subject of this dissertation (Jern *et al.*, 2005). Recent analyses show that more than twenty families of HERV exist. HERV sequence has been divided into three classes I, II and III (Hanke *et al.*, 2016) .These are classified according to sequence of their PBS, determining which tRNA they use for the priming of genome replication, for example HERV -K elements use lysine (amino acid single letter code K) and HERV-M elements use methionine (M) (Katoh & Kurata, 2013; Singh *et al.*, 2009).

Class I HERVs are most closely related to Gammmaretroviruses for example MLV. Class III HERVs are most closely related to Spumaviruses and are further divided into the HERV-L and HERV-S families (Gifford & Tristem, 2003).

As the HERV-K family of class II HERVs are the youngest and most recently active elements in humans, their subfamily structure will be considered in detail. HERV -K includes in class II, as illustrated in Figure (1.6).

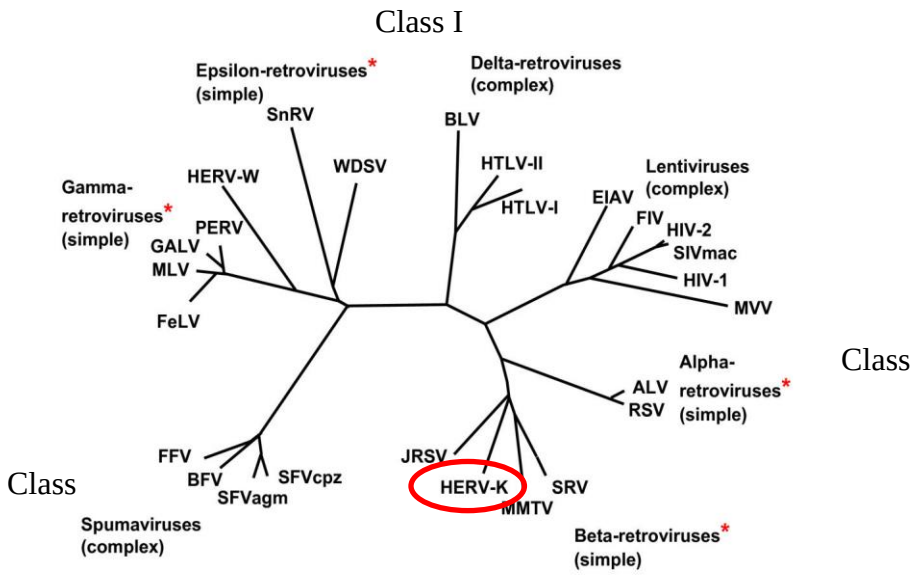


Figure 1.6: The classification of retroviruses

Phylogenetic tree of retroviruses showing the three classes of ERV and their relationships to exogenous viruses. HERV-K is highlighted within the Class II HERVs (red oval). Asterisks indicate the family of retrovirus (Weiss, 2006).

Studies have estimated that HERVs-K are represented by 550 proviruses and 6400 solo LTRs in the human genome (Hohn *et al.*, 2013). The HERV-K HML2 (hence the HML referred to human MMTV like. It is the youngest family and likely has significant biological activity because many elements have intact ORFs. As a result, some sequences of this family can produce viral particles, observed in the cytoplasm of cells (Suntsova *et al.*, 2015). In addition, some HERV-K HML2 insertions show presence or absence of polymorphism in humans, raising the possibility that genetic variation in HERV constitution and their biological activity may lead to differences in disease susceptibility or progression. In contrast, the class I and class III HERVs are much older

families that are not known to be active, and are fixed in human populations (Cegolon *et al.*, 2013).

The family of HERV-K (HML-2) retroelements can be classified into two types of proviruses, type 1 and 2, which can encode additional accessory, proteins (the nonstructural proteins *np9* and *rec*). These proteins are often expressed in transformed cells and tumors cells (Hohn *et al.*, 2013) . As result these proteins could play a role in cancer development (Singh *et al.*, 2009; Kassiotis, 2014).Type 1 proviruses encodes *np9* (a 9kDa protein) and have a 292-bp deletion at the *pol* and *env* boundary, resulting in disruption of the *env* gene (de Parseval & Heidmann, 2005). Although, the function of *np9*, in HERV-K replication is unknown, it is expressed in some cancer tissues (Singh *et al.*, 2009). Recently, some studies have suggested that *np9* is an important factor in human leukemia by altering the development of lymphoblastic and myeloid cells (Hohn *et al.*, 2013). In contrast to type 1, type 2 proviruses encode the protein *rec* (14.5kDa). When *rec* is intact, it can act significantly in cell transformation (Singh *et al.*, 2009). For example, *rec* expression affects the intensity of rheumatoid arthritis (RA) symptoms (Hanke *et al.*, 2013).

np9 and *rec* proteins bind to promyelocytic leukemia zinc-finger protein (PLZF) and inhibit the transcriptional inhibiting function of PLZF. This situation can lead to the promotion of cellular proliferation, instead of apoptosis, and so could have a role in tumorigenesis (Denne *et al.*, 2007).

1.6 The role of HERV-K in evolution

ERVs are thought to have played a significant role in the evolution of mammalian genomes, contributing important protein coding genes through the process of exaption. For example, syncytin-1 protein is encoded by HERV-W *env* during pregnancy to

develop the placenta. Moreover, HERV-W *env* induced immunosuppression is involved in protection of the fetus from mother's immune system. Also, HERVs have great potential as genetic markers for studying host evolution by using integration sites of HERV, under the assumption that they are very rarely deleted entirely once inserted (Kazazian, 2004). For example, human and baboons have HERV-W family insertions at the same orthologous loci despite being separated for over 30 Myr and such insertions can be used to date speciation events (Best *et al.*, 1997).

HERVs have integrated into the human germline between10-60 Myr, therefore they could affect host evolution by altering host gene expression or gene function. This is attributed to the transcriptional regulatory elements that reside within HERV LTRs (Lavrentieva *et al.*, 1999). Indeed, HERVs have inserted in germlines throughout vertebrate evolution: HERV-L proviruses are found in cow, mouse and dog genomes making these insertions over 100Myr old. HERV-I insertions are estimated to be around 400 Myr old, occurring in fish and birds (de Parseval & Heidmann, 2005). The HERV-K (HML-2) family integrated into the primate genome after the divergence catarrhine and platyrrhine primates (Figure 1.7). There is evidence for ongoing proliferation of HERV-K (HML-2) sequences in the great apes (orang-utan, gorilla, chimpanzee humans) as shown by curved arrows in figure 1.7. As a result, polymorphic HERV insertions are still present in the human population, with some individuals having the insertion; but with other individuals carrying only the empty pre-insertion site (Figure 1.8). Due to this polymorphism HERV-K elements have been investigated more than others LTR retroelements found in human DNA (Ilina *et al.*, 2012; Tristem, 2000).

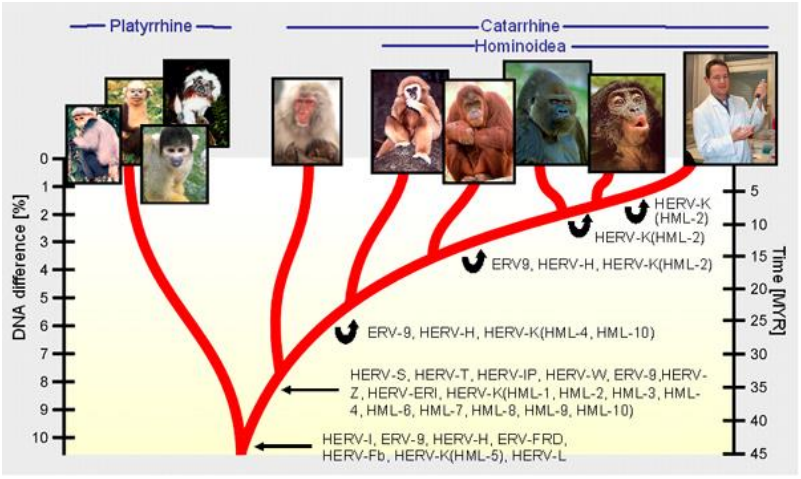


Figure 1.7: HERV insertion during primate evolution

Ongoing insertion for individual HERV subclasses is indicated with black curved arrows. The numbers on the left axis refer to the percentage difference in the genomic DNA of primates, compared to human, while the right axis refers to number of millions of years (MYR) from the diverged ancestor. The HERV-K (HML-2) family is present in the genomes of gibbons, orangutans, gorillas, chimpanzees and humans. It is suggested that HERV-K (HML-2) integrated more than 30 million years ago (from Leib-Mösch, C., 2010).

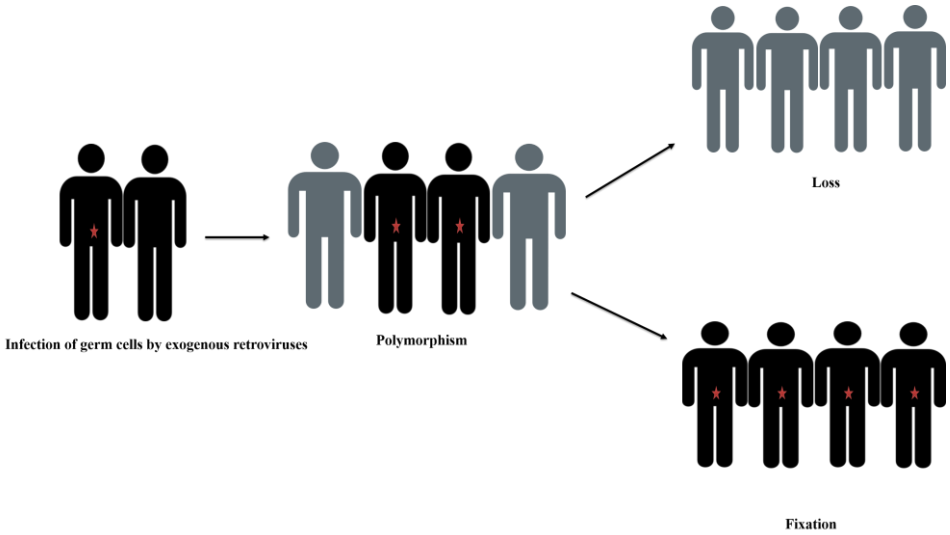


Figure 1.8: Hypothetical model for population spread of ERVs

Some endogenous RV insertions are lost from the population by drift, but others (black) are inherited as segregating polymorphisms. As the provirus accumulates mutations to become non-infectious and can only be transmitted vertically. The polymorphic ERV insertion may be lost (grey) or fixed in the host population.

1.7 Expression

In spite of the fact that exogenous and endogenous retroviruses are similar in structure, ERVs are simpler and defective due to deletion, mutation and recombination. As a result, we might expect that HERV-K, despite being the youngest family in humans, is unable to exert much influence on the host (Hohn *et al.*, 2013). However, HERV-K may remain a pathogenic gene because HERV-K proteins and transcriptional activity are associated with a variety of disease such as autoimmune disorders and cancers (Katoh & Kurata, 2013). In addition, homologous and non-homologous recombination of ERV can produce some mutations. Thus, ERVs contribute to instability of genome and could involve in initiation and development (Strissel *et al.*, 2012).

Dewannieux *et al.*, (2006) demonstrated that constructs derived from HERV-K provirus consensus sequences can integrate into the genome by a reverse transcription process. Similarly, they found that the in vitro recombination of HERV-K loci can behave like functional HERV-K elements. Thus, human cells still have the ability to express functional HERV-K viruses. It seems that whether HERVs are intact or not, they can impact the genome through recombination processes, by using two or three HERV-K defective copies to regenerate functional proviruses in human cells (Young *et al.*, 2013).

In rare cases, ERV genes have been expected to perform useful functions for the host. For example, the *env* gene of HERV-W elements has contributed sequences to the evolution of the Syncytin gene that has an essential role in placental syncytiotrophoblast development (Katoh & Kurata, 2013). Furthermore, genetic analysis has found that in mice, the Friend-virus-susceptibility-1 (Fv1) gene is derived from an ERV *gag* gene and protects against MLV infection (Katoh & Kurata, 2013).

With respect to younger HERV families, Kammerer *et al.* (2011) claimed that the HERV-K *env* gene is involved in the development of the placenta and pregnancy in humans. They found HERV-K TM protein was expressed in normal human placental tissue especially in cytotrophoblast cells. Since the HERV-K TM protein has immunosuppressive properties, expression of HERV-K TM protein may act in immune protection of the fetus.

It is also well established that the protein expression of sycytin from *env* of HERV-W has essential in development of placenta and protection of the fetus because of its immunosuppressive properties. In contrast, the abnormalities of placenta is related with the lost expression of syncytin protein of HERV-W (Suntsova *et al.*, 2015). Additionally, ERV-K *env* mRNA and proteins are expressed in normal ovarian tissue (Cegolon *et al.*, 2013). Ultimately, HERV genes may have other roles in human tissue development that we are unaware of (Katoh & Kurata, 2013).

Finally, HERV-K *env* has ability to inhibit tetherin activation which act as cell restriction factors and prevent binding any viral infection on cell surface. Thus, Tetherin is involved in innate immune responses (Nadeau *et al.*, 2015; Lemaitre *et al.*, 2014).

Recently, Grow *et al.* (2015) investigated the observation that in human embryogenesis HERV-K *gag* is expressed particularly at the eight-cell stage. In addition, the HERV-K *rec* (accessory protein) is highly expressed in pluripotent cells, which induces Interferon-induced transmembrane protein 1 (IFITM1) gene expression. IFITM1 is localized to the surface of cell and prevents binding of exogenous viruses in cell culture (Figure 1.9).

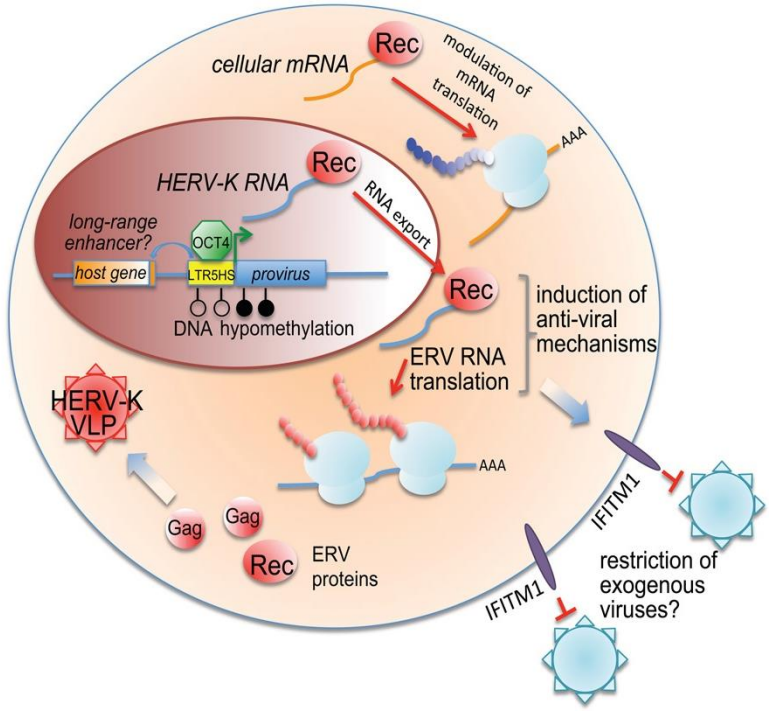


Figure 1.9: Model of HERV-K rec process

This mechanism shows the HERV-K *rec* acts on nuclear export of viral RNAs from nucleus to cytoplasm and may lead to induction of innate anti-viral responses, by causing expression of Interferon-induced transmembrane protein 1 (IFITM1), from (Grow *et al.*, 2015).

It has been noted that HERV-K *env* expression increases the expression of 300 genes such as IL-10 and decreased of another 300 gene such as SEPP1. Thus, it suggested that HERV-K *env* plays an important role in changing of all these genes and influencing on stability of human genes expression (Morozov *et al.*, 2013). In general, therefore, it seems that HERV-K expression impacts on genome expression by providing a new *cis* locus to start transcription (Suntsova *et al.*, 2015).

1.8 HERV-K and disease

HERVs can be expressed in human cells and be associated with different diseases due to many factors, as illustrated in Figure (1.10).

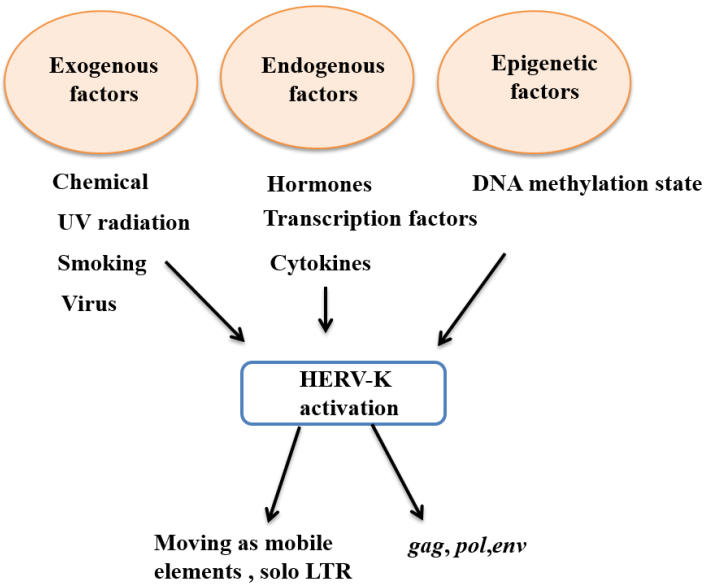


Figure 1.10: Schematic diagram of exogenous and endogenous factors

HERV expression is promoted by different factors, and may be associated with the development of human disease.

These effects are attributed to HERV RNA or protein expression, as well as the silencing of host genes (Lower, 1999). A number of research groups have proposed a causative role for HERVs in autoimmune diseases, for example, rheumatoid arthritis (RA) (Serafino *et al.*, 2009) and multiple sclerosis (MS) (Cegolon *et al.*, 2013). Recent work has shown that HERV-K *gag* expression is associated with RA, more frequently than in healthy groups (Hohn *et al.*, 2013).

Although not conclusive, it has also been proposed that HERV expression may serves as a risk factor for MS (Sicat *et al.*, 2005). However, Moyes's group in 2007

demonstrated that HERV-K113 and HERV-K115 expression is not associated with MS, indicating the relationship is far from established.

Furthermore, HERV-K RNA was expressed in blood of lymphoma patient without treatment compared with patient who was treated with different types of therapy such as cyclophosphamide, adriamycin, vincristine, and prednisone, indicating a low expression of HERV-K expression (Nadeau *et al.*, 2015). In leukemia patients, HERV-K *np9* protein was significantly increased in expression, relative to healthy individuals (Chen *et al.*, 2013).

In Morphea disease (autoimmune connective tissue) patients, HERV-K10 *gag* mRNA was increased after UVA1 phototherapy treatment, compared to healthy individuals. However, they did not record any statistical differences in HERV-K10 *gag* expression in Morphea patients before and after phototherapy treatment. Therefore, the effect of UVA1 phototherapy treatment on the HERV-K10 *gag* expression was not statistically confirmed (Kowalczyk *et al.*, 2017).

1.9 HERV-K and cancer

As well as autoimmune diseases, HERV-K activity may also be involved in a variety of cancers such as breast, ovarian, melanoma and prostate. This association may be connected with their ability to form Virus-Like Particle (VLPs) in cancer cells as illustrated in Table 1 (Cegolon *et al.*, 2013; Hancks & Kazazian, 2012).

According to Cegolon *et al*. (2013), HERV proteins are expressed at a high level in approximately 90% of epithelial ovarian cancer cases whereas, normal epithelial tissue only shows low levels of HERV protein expression (Lisanti *et al.*, 2013).

In the same way, Wang-Johanning *et al*. (2007) demonstrated that ovarian cancer patients have antibodies against HERV-K proteins in their sera, compared to healthy individuals. The presence of antibodies provides strong evidence of HERV-K protein expression. Recently some studies have argued that hypomethylation of the HERV LTR stimulates its transcription and may play a causal role in urothelial cancer, however hypomethylation of HERV-K has not been noticed in renal cancer (Cegolon *et al.*, 2013; Lavrentieva *et al.*, 1999). Earlier studies have noted that HERV-K *env* RNA occurs in 66% of breast cancers, in a mouse model (Cegolon *et al.*, 2013). Also, HERV-K *env* was investigated in patients who had breast cancer compared with healthy individuals and they found the migration of breast cancer cell was reduced HERV-K *env* after using siRNA to knockdown its (Zhou *et al.*, 2016).

Some studies observed that HERV-K *env* can act as biomarker in breast, prostate and lymphoma cancer (Nadeau *et al.*, 2015; Zhou *et al.*, 2016; Wallace *et al.*, 2014). In addition, Zhou *et al*. (2016) showed that the knockdown of *env* expression with shRNA*env* in breast cancer cells was observed to significantly decrease invasion and migration in mouse models and decreased cancer growth and size compared with shRNA control cells. Thus, HERV-K *env* could be used as a novel therapeutic target in these cancers. A recent study by Agoni *et al.* (2013), HERV-K *env* and solo LTR expression were detected in prostate cancer. Moreover, at a specific locus of (Chr 22q) two copies of HERV-K5 were found to be expressed in prostate cancer patients. However, the knockdown of HERV-K5, at the same locus, did not impact viability or apoptosis in a prostate cancer cell line (Goering *et al.*, 2015).

It has been noted that HERV-K VLPs (viral like particles) are expressed in pluripotent embryomal carcinoma and teratocarcinoma (NCCIT), and that their frequency was increased by induction of HIV-1 protein expression (Contreras-Galindo *et al.*, 2015).

In 2017, Lemaitre *et al.* published a paper in which they described that the HERV-K *env* protein changed the expression of the Endodermal to Mesenchyme Transition (EMT) in breast cancer cells. Thus, the cells became more mesenchymal in phenotype, and increased their motility. Also, they found by q PCR technique, HERV-K *env* influenced by transcription factors EGR1, ETV4, ETV5 and FosB expression, and theses transcription factors have been associated with EMT. Moreover, HERV-K *env* expression activated ERK1/2 signalling in breast cancer cell.

Some studies have used microarrays to study the influence of endogenous and exogenous factors on the expression of HERVs, for example UV exposure, chemical exposure, smoking and hormone levels (Muster *et al.*, 2003; Wang-Johanning *et al.*, 2007). For example, Hohenadl *et al.* (1999) showed that HERV-K *pol* expression is activated by UV exposure in primary epidermal keratinocytes of human skin. They investigated *pol* expression by using PCR products, and hybridized in reverse dot blot hybridization assay endogenous retroviral *pol* sequences. One question is whether the human genome reference sequence represents all the loci harboring HERV-K (HML-2) insertions that may have roles in other human disease.

Table 1. 1 A list of different cancers related with HERV-K expression in humans.

(Hohn *et al.*, 2013).

Tissue	Cancer	HERV-K(HML-2) activity	Reference
Skin	Melanoma	Retroviral particles Enhanced transcription RT activity Expression of Env, Rec, Np9	Buscher et al. (6), Muster et al. (86), Hirschl et al. (124)
Testes	Germ cell tumors, gonadoblastoma, seminoma	Anti-Gag/Env-Ab Expression of Rec, Np9	Boller et al. (87), Kleiman et al. (92), Boller et al. (125)
Ovary	Ovarian clear cell carcinoma; ovarian epithelial tumors	Expression of Gag and Env	Gotzinger et al. (15), Wang-Johanning et al. (72), Iramaneerat et al. (126)
Breast	Breast cancer	Free viral RNA RT activity Virus particles Specific CTLs	Wang-Johanning et al. (112), Contreras-Galindo et al. (127), Wang-Johanning et al. (128)
Prostate	Prostate cancer	Enhanced Gag-production due to fusion to androgen-dependent ETV1 and ETS genes	Tomlins et al. (12), Lamprecht et al. (13), Ishida et al. (111)
Blood	Lymphoma	Free RNA RT activity Virus-like particles	Contreras-Galindo et al. (127)

1.10 Mechanism of HERV K activation LTR in human cells

HERVs are expressed in different tissues (Chuong *et al.*, 2017) However, the control of this expression depends on the LTR promoter which can be bound by transcription factors (TF), resulting in the initiation of viral RNA synthesis. The LTR includes a range of TF binding sites (TFBS) and a TATA box, as illustrated in Figure 1.11 (Urnovitz & Murphy, 1996; Hu *et al.*, 2017).

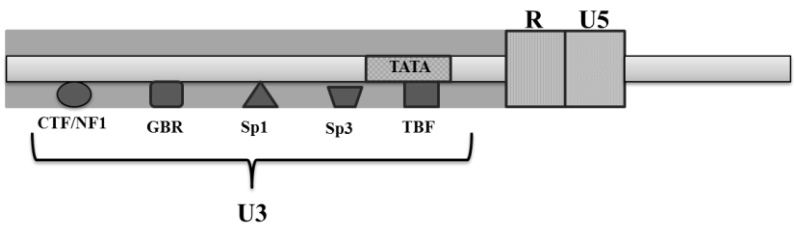


Figure 1.11 Diagram of HERV-K LTR structure

The LTR includes U3, R and U5 regions, of which the U3 region includes multiple TFBS of which Sp1 and Sp3 are the most important. (Diagram modified from Manghera & Douville, 2013).

HERV-K shows high expression in different diseases because Sp1 and Sp3 bind the LTR region, resulting in increased transcription. In support of this Fuchs *et al.*, 2011, demonstrated that knockdown of Sp1 and Sp3 by siRNA, resulted in low LTR activity.

It has be shown that TFs binding sites in the HERV-K LTR region influence neighboring gene expression and regulation. At the human amylase loci, amylase expressed in pancreas derived from two genes, while in salivary gland from three genes. All these genes contain a full length HERV-E upstream of their transcription start sites. In this situation, HERV-E is proposed to act as an enhancer to a cryptic promoter for increased transcription of the amylase genes (Suntsova *et al.*, 2015; Manghera & Douville, 2013)

Generally, Suntsova *et al.* (2015) and Thompson *et al.* (2016) collated in their publications that the HERV-K LTR can impact gDNA expression by a range of mechanisms, summarized below (Figure 1.12.)

1. Enhancer sequences inserted upstream of the LTR may induce transcription because the LTR includes many TFBS.
2. The LTR can act as a new promoter to drive transcription: nearly 114 transcription start sites of HERV LTR can drive 97 human genes' expression. However, the activation of the LTR is influenced by the type of LTR, such as 5′LTR, 3′LTR and solo LTR. In spite of the fact that the 5′LTR and 3′LTR are identical in sequence, the 5′LTR is observed more highly active than 3′LTR.
3. The premature signalling of polyadenylation could disrupt transcription and or translation of genes. For example, a polyadenylation signal in a HERV 5′LTR terminates transcription of the ZNF195 gene (Kjellman *et al.*, 1999).

4. Sometimes LTR insertion can occur in intronic regions of gDNA, and may damage exon –intron structure because the LTR causes abnormal splicing in the structure of genes. For example the Hpr and HP genes in human are 92% identical and both are expressed in liver. However, the HP transcripts is 17 fold higher than Hpr gene. The main cause is that the intron region of Hpr gene includes an RTVL-Ia provirus. Thus RTVL-Ia may contribute to change in the intron-exon region and abnormal splicing of Hpr transcript.
5. Insertion of an LTR in the antisense direction relative to a gene may produce complementary mRNA, which can form double stranded RNA that is degraded.

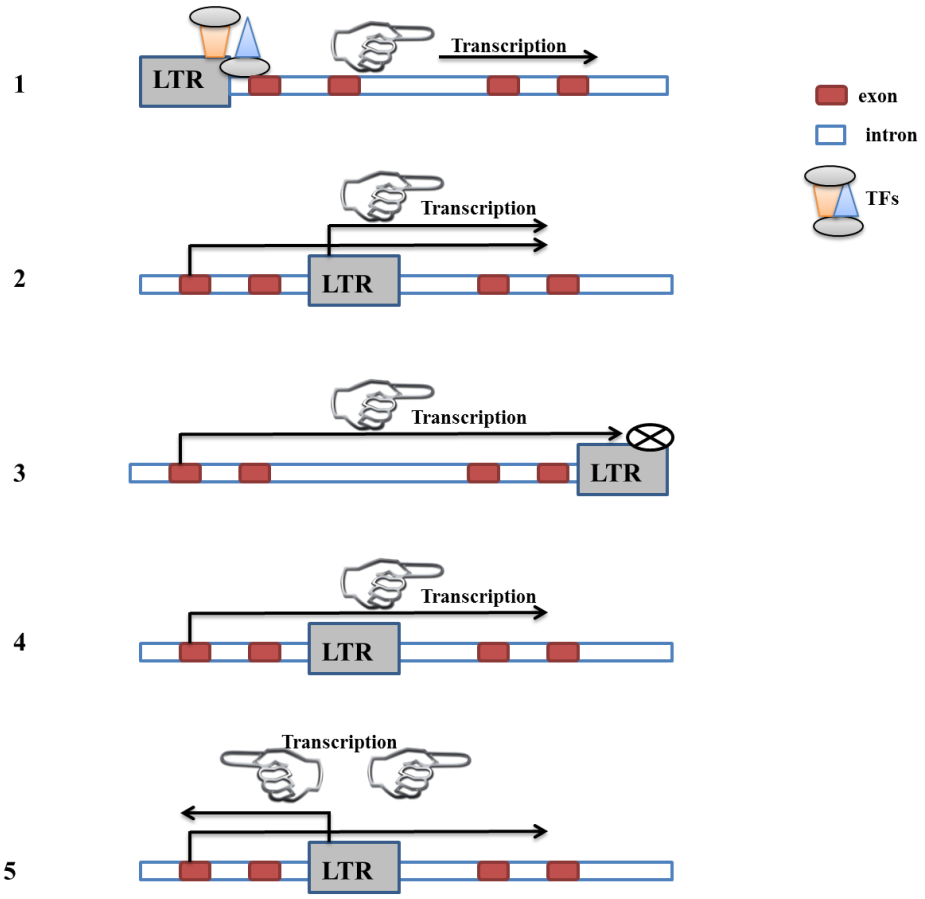


Figure 1.12: Mechanisms LTRs interact with gene expression

Five process of LTR interaction that influence gene expression, according to their site of insertion.

1.11 Melanoma

Melanoma is the most aggressive type of skin cancer and is recorded a top 10 of cancer types according to mortality rate. Melanoma is considered to be the most highly invasive type of skin cancer as it becomes metastatic rapidly and because of its resistance to many chemo and radiation therapies (Timar *et al.*, 2016).

Melanoma arises from melanocytes, which produce the melanin pigment in the skin, this pigment is responsible for skin color and its expression is regulated by epidermal keratinocytes (Gray-Schopfer *et al.*, 2007). There are many risk factors that lead to the development of melanoma from melanocytes, but the primary factor is UV light exposure (Timar *et al.*, 2016; Riverso *et al.*, 2017). UV exposure has changed C to T base in DNA sequence, nearly 75% of UV mutated occurs in BRAF gene in exon 15 codon600 and second UV mutated is in RAS gene in exon 3 codon 61, as result both of these mutations can play an important role for changing many signal pathways such as RAS/BRAF/MEK/ERK signal pathway (Timar *et al.*, 2016).

Another major risk factor is family history: each member of a family diagnosed with melanoma doubles the chance of development of melanoma. The common two mutations in familial melanoma are cyclin-dependent kinase Inhibitor 2A (CDKN2A) and cyclin depended kinase (CDK4), nearly 5%–10% of melanoma that occur by these mutations. A recent population-based. Therefore, the individual who has family history will be at increased risk of melanoma because of the mutations above (CDK4) (Box *et al.*, 2001; Kohli *et al.*, 2017).

In addition to family history, epigenetic state can play an important role in the development of melanoma. Pergoli *et al.*, (2014) showed that the hypomethylation of the telomerase reverse transcriptase gene (hTERT) and Alu elements were associated

with development of melanoma. Also, they did not find a correlation between global methylation of L1 elements and development of melanoma.

1.12 Stages of melanoma

Gray-Schopfer *et al.* (2007) explored the progression of melanoma from normal melanocytes to malignant melanoma, as illustrated in Figure 1.13. Many factors contribute to the causes of melanoma, from melanocyte mutation, expression of autocrine growth factors and the loss of cell adhesion receptors. Generally, melanoma progression is histologically separated into five stages: At stage one, a naevus, or common mole results from uncontrolled growth of melanocytes. The second stage is known as the junctional naevus stage, where melanocyte growth is limited to the epidermal layer. Later, the dermal naevus stage occurs which appears to be benign. However, it is an unstable stage because the cells could develop to invade the dermis layer and then enter the radial-growth-phase (RGP).

Cells at the RGP stage can develop to the vertical-growth phase (VGP). Interestingly, it is at the VGP stage where the highest risk of metastasis occurs where melanoma cells develop the ability to migrate through the bloodstream to any place in the body, but especially the brain and lymph nodes. Melanoma stages are either classified according to the Clark model, which depends on the depth of lesion and the skin layers affected, or the Breslow model, which uses the thickness of melanoma in the skin as an indicator of the stage of the cancer. Moreover, melanoma can be classified into four common subtypes: nodular melanoma, acral lentiginous melanoma, lentigo maligna and superficial spreading melanoma (Liu & Sheikh, 2014).

It has been demonstrated that the different subtypes of melanoma can be distinguished by their somatic mutations, with 70% accuracy (Gray-Schopfer *et al.*, 2007).

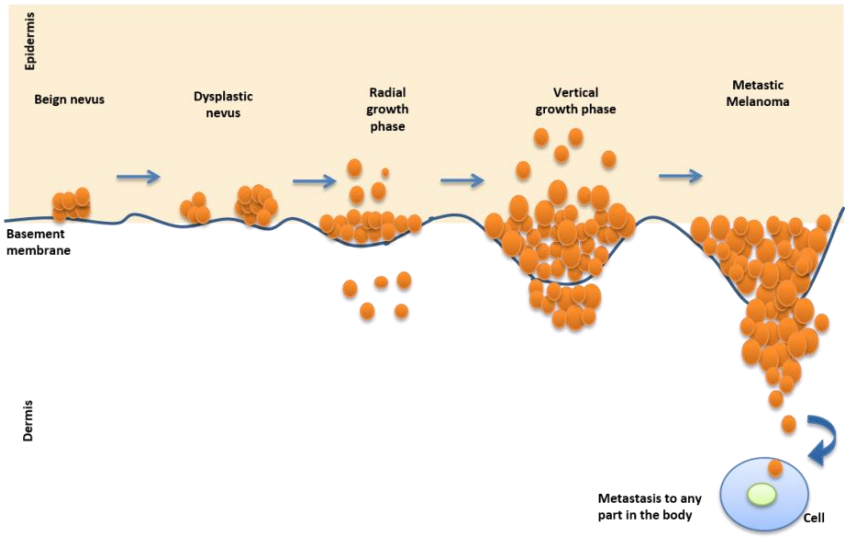


Figure 1.13: Biological processes in melanoma

Five stages of melanoma, which is produced from melanocytes. The development starts at a benign nevus, and progresses to the metastatic stage, whereupon the melanoma can spread to any part of the body.

1.13 Signaling pathways in melanoma

Human melanoma demonstrates a wide range of genetic changes that increase cellular proliferation and reduce apoptosis and induce epithelial mesenchymal transition. The most commonly mutated signaling pathways are BRAF or RAS signaling, and E-cadherin signaling.

RAS signaling is regulated in melanoma cell proliferation by RAS/RAF/MEK/ERK, or MAPK (mitogen activated protein kinase), and the RAS/PI3K/AKT signaling pathways. The BRAF /MAPK pathway plays a vital role in regulating melanocyte proliferation (Figure1.14).

The most frequent mutation is that of BRAF, observed in 40-60% of melanomas, (Muñoz-Couselo, E., García, J., Pérez-García, J., Cebrián, V., Castán, J., 2015).

One particular BRAF mutation (specifically V600E BRAF) that has been extensively studied, occurred in an individual who had melanoma arising from skin with sporadic sun exposure, and not in skin with long time sun exposure (Liu & Sheikh, 2014). However, this mutation is also found in benign nevi and in primary and metastatic melanomas with the same frequency (Pollock *et al*. 2003). The V600E BRAF mutation leads to constitutive activation of ERK (extracellular signal-regulated kinase) signaling, which is essential for cell survival. Thus, it is recommended that MEK inhibitor is used instead of V600EBRAF inhibitor to treat melanoma or reduce it, because of its resistance to immunotherapy of melanoma treatment (Bu & Diehl, 2016).

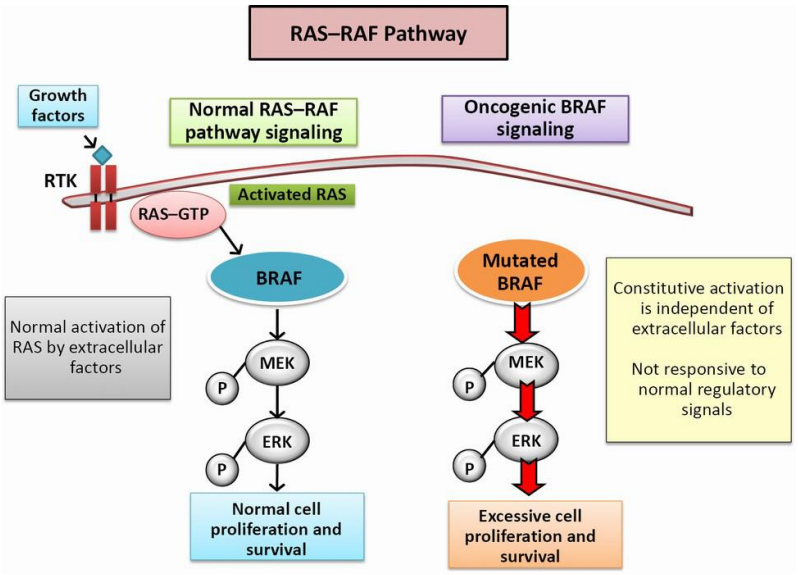


Figure 1.14: MAPK pathway in melanoma

This pathway includes ERK, MAP, and MAPK signaling. Normal BRAF (blue oval) signaling is dependent upon extracellular factors line. When BRAF is mutated (orange oval) signaling is constitutive, leading to melanoma (Muñoz-Couselo, E., García, J., Pérez-García, J., Cebrián, V., Castán, J., 2015).

E-Cadherin is present in polarized epithelial cells in the epidermis, including melanocytes and keratinocytes. E-cadherin is an important adhesion factor between keratinocytes and melanocytes in the epidermis. In melanoma, E –cadherin is not expressed, rather N-cadherin is expressed in melanoma cells. N-cadherin expression allows them to change phenotype from epithelial to mesenchymal. Thus, N-cadherin provides motility and invasion in melanoma, (Bonnelykke-Behrndtz *et al.*, 2017). It has been noted that transfer from the radial-growth stage to the vertical-growth stage of melanoma is the result of losing E-cadherin expression, and subsequent expression of N-cadherin (Liu & Sheikh, 2014). Similarly, integrin is a cell surface adhesion factor which plays an important role in cell adhesion, proliferation, migration, and apoptosis in normal tissue. However, in melanoma integrin is expressed at low level and may play an important role in migration, and survival in melanoma cells (Bonnelykke-Behrndtz *et al.*, 2017).

1.14 HERV-K (HML-2) and Melanoma

The most malignant kind of skin cancer is melanoma. Patients with aggressive malignant melanoma can die 6-8 months after diagnosis (Huang *et al.*, 2013).The expression of HERV-K proteins has been investigated in primary melanoma and metastatic melanoma cells, but HERV-K expression in melanocytic lesions is not evident (Serafino *et al.*, 2009). This may be because HERV-K activity is associated with the early developmental stage of melanoma cells (Muster *et al.*, 2003). Moreover, one study demonstrated that HERVs-K retrovirus-like particles (RVLP) are found in cells and tissue of human melanoma, and that their expression changes the morphology and growth phenotype of melanoma cells by transforming them from adherent to non-adherent cells, at least in cell culture (Serafino *et al.*, 2009).

Although, HERV-K (HML2) proviruses lack infectivity, many studies have mentioned that the HERV-K (HML2) plays a vital role in melanoma associated retrovirus (MelARV) infection, by increasing metastatic activity especially in primary B16 melanoma cell lines (Singh *et al.*, 2009).

Recently, some studies have investigated whether HERV-K elements play an important role in human melanomas. For example, HERV-K mRNA is expressed in 50% of melanoma cells lines and metastatic melanoma tissue, and HERV-KTM and HERV-K gag protein are observed in metastatic and primary melanoma cell with 80% and 50% frequency, respectively (Buscher *et al.*, 2005; Muster *et al.*, 2003).

The HERV-K108 *pol* protein (RVLP) is also recorded at significantly higher rates in melanoma patients tissue compared to healthy people by in *situ hybridization* (Singh *et al.*, 2009). In addition, HERV-K *gag* and *env* antibodies occur in the sera of 16% (51/312) of melanoma patients in contrast to healthy individual (Cegolon *et al.*, 2013).

The American Joint Committee on Cancer (AJCC) noted a high prevalence of antibodies against HERV-K *env* proteins in 81 melanoma patients with AJCC stages I-III disease while 95 sera healthy individuals lacked antibodies (Cegolon *et al.*, 2013). Buscher *et al.* (2005) showed that the *rec* protein and spliced *env* antisera are detected in 39% of metastatic melanoma tissue samples and 40% of cell lines. However, the *np9* protein has not been detected in melanoma tissues by immunohistochemistry while TM and *rec* proteins are recorded in 38% and14% of cases, respectively (Buscher *et al.*, 2006). It has been reported that HERV-K *rec* and *np9* proteins expression is induced by smoking, hormones (androgen) and UV radiation, as environment factors, and may be related to cause or development melanoma (Hanke *et al.*, 2016)

Krishnamurthy *et al*. (2015) found that HERV-K *env* CAR (chimeric antigen receptor) can be used as a good target to inhibit melanoma cells expressing HERV-K. Although, this experiment was done in mouse model, it is a new therapeutic process that could be applied in humans. It has been found that HERV-K *gag* protein was detected in melanoma cells, and p-ERK protein was associated with expression of *gag*, while P16INK4A was reduced with increased expression of *gag* protein (Li *et al.*, 2010).

Recently, Argaw-Denboba *et al.* (2017) reported the effect of serum starvation on the TVM-A12 human melanoma cell line. This cell line was cultured in normal conditions (10% FBS) and also in a serum free stem cell medium (X-VIVO-15). The phenotype of cells in serum-free media changed to non-adherent under these stress condition due to reduced expression of cell surface markers such as HLA-I, MelanA/MART-1 and ICAM-1. In addition, this study showed the HERV-K expression under starvation culture was increased and associated with a CD133+ subpopulation expression. Also, they investigated the down-regulation of HERV-K expression which was associated with the reduction of CD133 expression, on treatment with reverse transcriptase inhibitors (NNRTIs).

Taken together, these results demonstrate that TVMA12 human melanoma cells have a change-prone phenotype in response to microenvironment modification, characterized by the expansion of a CD133+ (stem cell marker) melanoma subpopulation, enhanced expression of HERV-K, and of markers related to immune evasion and metastasis formation.

In conclusion, HERV-K may be postulated to have some causal role in melanoma, perhaps through modulating the immune system's activity in disease, but the exact role of HERV-K in melanoma has not been defined (Singh *et al*., 2009).

1.15 Epigenetic modifications

All somatic cells in eukaryotes generally contain the same genome, but in metazoans each cell type has a different function and structure because of variation in gene expression between cells. The way this variation is created and maintained is not fully understood but one type of explanation is that different cell types have different patterns of reversible regulatory information laid down upon their DNA that can be inherited intact through cell division without change to the underlying DNA sequence. This process is called epigenetics, and has been defined as "any changes of heritable information in gene expression which occur without any change of DNA sequence" (Menendez *et al.*, 2004).

Known epigenetic modifications include DNA methylation and chromatin modification. These modifications are achieved by DNA methyltransferases (DNMTs), histone demethylases (HDMTs), histone methyltransferases (HMTs), histone deacetylases (HDACs) and histone acetyltransferases (HATs) enzymes whose activities on DNA and histones, influence gene expression (Wolffe & Matzke, 1999). It has been shown that epigenetics can play an important role in differentiation and development, as well as pathogenesis (Bestor, 2000). Some studies have suggested that the relationship between epigenetic modification and diseases such as cancer is strong. Some research mentioned that DNA hypomethylation is associated with progressing or increasing tumor in tissue compared with normal tissue which is hypermethylated level (Feinberg & Vogelstein, 1983; Ehrlich, 2002). For example, HERV-W and L1 are hypomethylated in ovarian cancer, in contrast to normal ovarian tissue (Menendez *et al.*, 2004). In addition, in germ cell cancer, the hypomethylation of CpG regions upregulated HERV-K expression. Thus, CpG methylation level contributes to transcriptional activity of HERV-K LTR and increased expression of HERV-K (Lavie *et al.*, 2005).

1.15.1 DNA methylation

In vertebrate genomes between 70-80% of cytosine bases in the CpG dinucleotide are methylated by addition of a methyl group (CH_3) to the 5' carbon of cytosine bases as illustrated in Figure (1.15). This phenomenon is called CpG dinucleotide methylation (Florl *et al.,* 1999). DNA methylation can be described according to the increase or decrease in the density of methylated cytosine compared to the genome average. When methylation is frequent, this is called hypermethylation and when infrequent it is hypomethylation (Menendez *et al.*, 2004). However, small regions of the genome about 1Kb long which are unmethylated in normal cells, but contain a high density of CpG dinucleotides are called CpG "islands". Approximately 15% of all CpGs in the human genome occurs within CpG islands (Sukapan *et al.*, 2014). CpG islands are often located in promoters or the first exon of genes, where they can affect transcription activity (Antequera & Bird, 1993). On the other hand, if the CpG island becomes methylated, gene expression will be repressed. CpG islands can be found methylated at four places in the human genome: the inactive X-chromosome in females, imprinted genes and genes showing tissue –specific and germline –specific expression. Some data also suggest that CpG islands become methylated with increasing age in humans (Sukapan *et al.*, 2014).

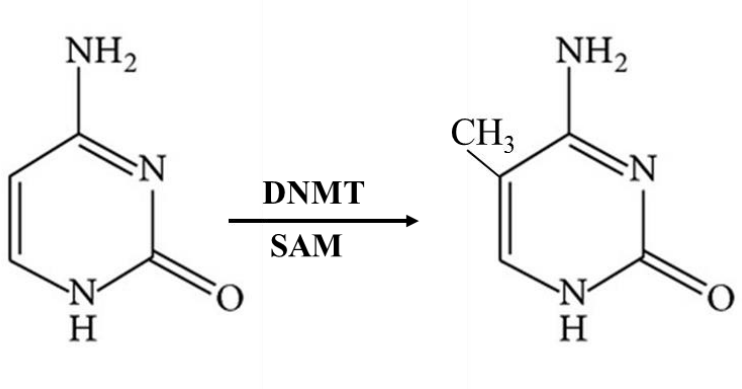


Figure 1.15: DNA methylation structure.

DNA methylation occurs by the covalent addition of a methyl (CH_3) group at the 5-carbon position of the cytosine, by enzymes called DNA methyltransferases (DNMT) and S-Adenosyl-methionine (SAM) , resulting in 5-methylcytosine. Approximately 2-3% of mammalian genomic DNA is 5-methylcytosine.

DNA modification is achieved by DNA methyltransferases that are classified into two types: maintenance DNA methyltransferases (for example DNMT1 in mammals) and *de novo* methylases (e.g. DNMT3A and DNMT3B in mammals). Maintenance DNA methyltransferases act on hemi-methylated DNA and the major type found in humans is encoded by the DNMT1 gene (Sukapan *et al.*, 2014). Furthermore, DNMT1and HDAC may repress transcriptional activity together (Yoder *et al.,* 1997). The human *de novo* methylases (DNMT3A, DNMT3B) are responsible for creating new patterns of DNA methylation (Bestor, 2000). Many studies have shown that retro-elements are usually hypermethylated, including endogenous retroviruses. It is established that DNA methylation may silence expression of retrotransposons and can control their expression and hence mobilization, in the genome (Yoder *et al.*, 1997). Hypomethylation of HERVs and other retrotransposons has been reported in cancer, ageing and autoimmune disease (Ehrlich, 2002).One study has shown that the HERV-K LTR has different levels of methylation in different tissues, for example the cerebellum versus lymph node tissue (Lavie *et al.*, 2005).

Hypomethylation level may induce some transcription factors that than increased expression of HERV-K in some cells lines. However, the most studies suggest that the variation in transcriptional activity in cells lines does not rely entirely on CpG methylation. For example, although CpG methylation silences HERV-K (HML-2) in normal cells, HERV-K (HML-2) is significantly expressed in the T47D cell line with hypermethylated level of CpG region (Lavie *et al*., 2005).

1.15.2 Chromatin modification

DNA is compacted in the cell using proteins, and this protein packaged form of DNA is called chromatin (Zhang & Reinberg, 2001). In addition, the nucleosome is a basic unit of chromatin and contains 8 subunits, 2 of each four core histones H3, H4, H2A and H2B around which 174 base pairs of DNA are wrapped (Liu *et al.*, 2011). All known modification of histones occurs at their N-terminal tails, which are unstructured and extend away from the DNA. The modifications are acetylation, methylation and phosphorylation which occur at specific N-tail amino acids, and often influence gene expression, most likely through the packaging density of the DNA (Strahl & Allis, 2000). Chromatin can be classified into two types: heterochromatin and euchromatin. Heterochromatin is a generally silent form of DNA in the genome, but is essential as a structure because it promotes separation of chromosomes, and the ends of chromosome during mitosis (Bernstein *et al.*, 2005). Accordingly, heterochromatin is associated with high levels of DNA methylation and low DNA acetylation levels in mammals, such as methylation of Lys 9 of histone H3 (H3K9), thus heterochromatin does not play an important role in transcription and gene expression (Zhang & Reinberg, 2001).

In contrast, euchromatin encompasses a high proportion of the genome (Baylin *et al.*, 2001). Euchromatin plays a vital role in biological activity such as replication of DNA, repair of DNA and transcription. For example, enzymes that are important for transcription occur in DNA with high methylation levels such as methylation of Lys 4 of histone H3 (H3K4) (Bernstein *et al.*, 2005).

More complex chromatin modifications occur by methylation of lysines and arginines in the histone N-terminal tails as illustrated in Table 2. These modifications can be found in one of three forms called, mono, di and tri (Liu *et al.*, 2011).

Table 1. 2 List of different classes of modifications found on histones.(Sadakierska-Chudy & Filip, 2015)

Type of modification	Amino acid (symbol)	Level of modification	Abbreviation of modifications	Examples of notion
Methylation	Arginine (R)	Mono-	me1	H3R2me1
	Arginine (R)	Di-, symmetrical	me2s	H2ARme2s
	Arginine (R)	Di-, asymmetrical	me2a	H4R3me2a
	Lysine (K)	Mono-	me1	H3K9me1
	Lysine (K)	Di-	me2	H3K9me2
	Lysine (K)	Tri-	me3	H3K9me3
Acetylation	Lysine (K)	Mono-	ac	H4K5ac
Phosphorylation	Serine (S)	Mono-	ph	H3S10ph
	Threonine (T)	Mono-	ph	H3T11ph
	Tyrosine (Y)	Mono-	ph	H3Y41ph
Ubiquitination	Lysine (K)	Mono-	ub1	H2AK119ub1
	Lysine (K)	Di-	ub2	H2AK119ub2
	Lysine (K)	Poly-	ubn	H2AK119ubn
Sumoylation	Lysine (K)	Mono-	su	H4K14su
Biotynylation	Lysine (K)	Mono-	bio	H2AK9bio
Citrullination	Arginine (R)	Mono-	cit	H3R17cit
ADP-ribosylation	Glutamate (E)	Mono-	ar1	H1E15ar1
	Arginine (R)	Mono-	ar1	H1.3R33ar1
	Glutamate (E)	Poly-	arn	H2BE2arn
β-N-glycosylation	Serine (S)	Mono-	glc	H3T32glc
	Threonine (T)	Mono-	glc	H3S10glc
Isomerization	Proline (P)	*cis/trans*	iso	H3P38iso
	Aspartic acid (D)		iso	H2BD25iso?
Crotonylation	Lysine (K)	Mono-	cr	H2BK5cr
Formylation	Lysine (K)	Mono-	fo	H1K17fo
Propionylation	Lysine (K)	Mono-	prop	H3K23prop
Butyrylation	Lysine (K)	Mono-	buty	H4K5buty

It has already been noted that DNA methylation is generally associated with inactive chromatin. (Ng & Bird, 1999) showed that the herpes simplex virus thymidine kinase (HSV TK) gene has the same rate of transcription from methylated and non-methylated DNA over 8 hours from microinjection in rodent cells, but after 8 hours the transcription activity from methylated DNA stops. The explanation that Ng & Bird presented was that initially DNA methylation did not impact on chromatin structure or transcription activity, but that after 8 hours the DNA methylation led to modified chromatin structure, and so resulted in transcriptional repression.

It is evident from numerous studies that gene expression often depends on chromatin modification and DNA methylation (Bestor, 2000). These modifications could be activating or repressing depending on the locus. For example, the histone mark H3K36me is activating in coding regions but repressing at promoters (Sukapan *et al.*, 2014). In general, however, three methylation sites on histones H3K4, H3K79 and H3K36 activate gene expression. On the other hand, methylation sites implicated in repression of gene expression occur on H3K27, H4K20 and H3K9.

1.15.3 Methyl Binding proteins in epigenetic modification

An important step in the epigenetic modifications within cancers is related with methyl-CpG binding domain (MBDs) containing proteins. These proteins play a vital role in hypermethylated gene promoters and transcriptional silencing. The mammalian family of MBD- containing proteins includes five members, MeCP2, MBD1, MBD2, MBD3 and MBD4. All MBD – containing proteins associate with histone deacetylases (HDACs) and through DNA methylation lead to silencing of the transcription process (Ballestar *et al.*, 2003). MeCP2 is a protein that binds methylated DNA and on binding is able recruit histone deacetylases because it can form a stable complex with Sin3A and

HDAC, preventing gene expression (Ehrlich, 2002). For example, in mice, methylation of the M-lysozyme gene in non-myeloid cells prevents gene expression in lymphocytic T cells and Myeloid precursor cells, but treatment with TSA (Trichostatin A: an inhibitor of HDAC) restores gene expression. Therefore, the link between DNA methylation and chromatin structure is very strong, but not completely understood (Ng & Bird, 1999).

1.16 Genome-wide hypomethylation

It has been noted that genetic alterations can cause different types of cancer by mutation, deletion and epigenetic modification, each of which change gene expression. Recent studies indicate that hypermethylation of tumor suppressor genes can be found in some human cancers, and promoter hypermethylation may play a critical role in carcinoma (Baylin *et al.*, 2001; Stengel *et al.*, 2010).Toyota *et al.*, (1999) suggested that the risk of colon cancer is increased significantly in ageing due to hypermethylation of gene promoters. On the other hand, genome-wide hypomethylation is a common phenomenon in multiple types of cancer (Toyota *et al.*, 1999). This point is emphasized by many studies showing that cancer tissues contains 2-3 % methylcytosine, compared to normal tissue which includes approximately 4% methylcytosine (Lisanti *et al.*, 2013). Ehrlich (2002) noted that this lower level of methylation in cancer cells occurs largely in repetitive elements such as LINE-1, HERV-K and Satellite 2. Several methods have been developed to measure the level of methylation in genomes such as HPLC (high performance liquid chromatography). Although, this method is an absolutely quantitative measure of methylation, it needs large amounts of DNA. Other methods use less DNA, such as bisulphite PCR and luminometric methylation assays (LUMA), but are less quantitative (Lisanti *et al.*, 2013).

Genome -wide hypomethylation in cancer can occur in two ways: passively, which means the loss of maintenance of methylation during cell division and actively which refers to the loss of cytosine methylation from DNA directly, without involving replication. Active DNA demethylation can occur by an enzymatic process involving the Ten eleven translocation (TET) family of dioxygenases. This family includes three types of enzymes, namely TET1, TET2 and TET3 that catalyze the successive oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC).The 5hmC oxidation products enable conversion of 5mC to unmodified cytosine (resulting in the loss of the methyl group), providing the active DNA demethylation pathway. As 5-hmC is enriched within gene promoters, and transcription factor binding sites, it may impact on gene expression (Alaghband *et al.*, 2016). Active demethylation occurs normally in mammalian genomes, especially in gametogenesis and early embryogenesis, but the enzymology is only just beginning to be understood (Sukapan *et al.*, 2014).

Several studies have demonstrated that hypomethylation at > 25% of CpG dinucleotides occurs in breast, ovarian, prostate and melanoma cancers (Menendez *et al.*, 2004; Baylin *et al.*, 2001). Hypomethylation of HERV-W DNA has been observed in human cancers, in contrast with normal cells (Stengel *et al.*, 2010). Furthermore, HERV hypomethylation is reported in prostate and bladder cancer (Lisanti *et al.*, 2013). As a result HERV methylation may be considered as a biomarker for cancer development (e.g. melanoma and breast cancer) (Kassiotis, 2014). In addition to HERV proviruses, solo LTR and MaLRs have been implicated in many types of carcinoma (Sukapan *et al*., 2014). Florl *et al.* (1999) described how L1 and HERV-K could be associated with urothelial cancer in hypomethylated conditions. Similarly, HERV-K is expressed when its promoter is hypomethylated in melanoma cell lines (Stengel *et al.*, 2010). Genome-wide hypomethylation may be associated with tumor initiation. For example, it has been

shown that HERV-K expression is increased substantially in prostate cancer, which correlates with hypomethylation at HERV loci (Katoh & Kurata, 2013).

Another point that has been some bearing on the situation is the fact that the hypomethylation that occurs during loss of imprinting (LOI) is related to tumor growth. It is observed that individuals who have LOI, are at increased risk of developing colon cancer (Ehrlich, 2002).

1.17 Tumor suppressor gene methylation

It is clear that epigenetic modification is affected by the interpretation of genetic information. Because DNA methylation plays an important role in the expression of cell type–specific genes, in some cases such as cancer, this may impact on cellular transformation (Baylin *et al.*, 2001).

Hypermethylation of tumor suppressor genes has been observed in many kinds of cancers (Sukapan *et al.*, 2014), and many studies indicated that DNA hypermethylation can silence tumor suppressor genes (Bernstein *et al.*, 2005). It has been shown that hypermethylation of CpG islands of tumor suppressor genes influences signaling cellular pathways; for example, CDKN2A hypermethylation affects the cell cycle (Esteller *et al.*, 2001). Furthermore, H3K9me silences tumor suppressor gene CDKN2B and CDH1 in myeloid leukemia cells (Baylin *et al.*, 2001). Another important point is that TP53 could be considered the most common mutation of a tumor suppressor gene implicated in human cancers. Hypomethylation has been observed to cause silencing of CDKN2B, and as a result TP53 can became inactivated, contributing to cancers such as breast, prostate and colon cancer (Baylin *et al.*, 2001).

Therefore, epigenetic modifications can play a vital role in tumor development by silencing tumor suppressor genes (Figure1.16)

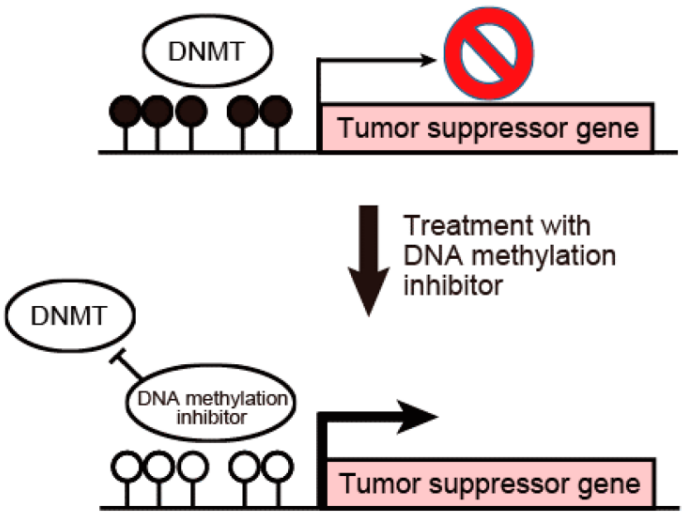


Figure 1.16: Activation and silent of tumor suppressor genes.

DNA methylation inhibits tumor suppressor gene promoter regions. In cancer cells, tumor suppressor genes are silenced by DNA hypermethylation on promoter regions. In principle 5aza dC DNA methylation inhibitors like 5-aza cytidine could reduce DNA methylation level in promoter regions. Dark circles indicate methylated DNA; white circles indicate unmethylated DNA. (From Nakaoka *et al.*, 2017)

1.18 LTR and epigenetics

Although, the HERV-Ks have been not been active in recent evolutionary time because of mutation, recombination and deletion, their epigenetic state plays an important role in silencing HERV-K activation in different tissues, by methylation CpG regions in the LTR sequence (Manghera & Douville, 2013).

The hypomethylated state is generally seen as a factor strongly related to cancer compared with healthy tissue. It is clear that activation of HERV-K with hypomethylation has the potential to contribute to cancer initiation or its development. However, it is difficult to decide about the positive correlation between HERV-K and

cancer because while many researcher found the HERV-K expression is associated with cancer (Lavie *et al.*, 2005; Serafino *et al.*, 2009; Schmitt *et al.*, 2013; Krishnamurthy *et al.*, 2015) many others did not find any such correlation (Chen *et al.*, 2013; Singh *et al.*, 2013; Hurst *et al.*, 2016).

What is clear is that the transcription activity of HERV-K depends on the methylation state of LTR, with a high level of methylation LTR region leading to low expression of HERV-K in tissues. Thus, this process can act as defense mechanism to silence HERV-K activity. For example, in the Tera -1 embryonic carcinoma cell line, expression of HERV-K with hypomethylation of LTR region was observed (Lavie *et al.*, 2005). However, SKMel13 (melanoma cell line) did not express of HERV-K with hypomethylation level of the LTR promoter (Stengel *et al.*, 2010; Romanish *et al.*, 2010). It is clear that the epigenetic (DNA methylation) state can regulate HERV-K expression in some contexts rather than always silencing it.

Histone modification also may influence the inhibition of HERV-K activation. Hurst & Magiorkinis (2017) demonstrated that the histone deacetylase inhibitor (HDACi) did not increase HERV-K expression in T cells infected with HIV1. This study confirmed that chromatin deacetylation is not the main cause for HERV activity.

Recently studies have investigated that double strand RNA (dsRNA) can be induced by using 5-aza-2'-deoxycytidine (5azdC), a DNA methylation inhibitor. The activation of dsRNA increases interferon gene expression, and this process can be used as an immune check point inhibitor in most types of cancer. Specifically, inhibition of methylation in HERV LTR regions in melanoma cells, meant they expressed dsRNA and became more sensitive to treatment with antibodies that target CTLA4 an inhibitor of the recognition of melanoma cells by the immune system (Chiappinelli *et al.*, 2015).

Concurrently, another study by (Roulois *et al.*, 2015) found that dsRNA from ERV was up regulated after treatment cancer –initiation cells (CIC) by 5azdC inhibitor. These authors also observed an increase in MDAS/MAVS signaling, which induces an interferon response, as illustrated in Figure 1.17.

Furthermore, the knockdown of TLR3 and MAVS (important signaling pathways in ovarian cancer) was down regulated two fold by 5azdC treatment and worked to induce an interferon response.

Other mobile elements, such as L1 can contribute with to disease in a methylation dependent manner. For example, Sukapan *et al*. (2014) found that L1 expression and hypomethylation level was involved in systemic lupus erythematosus (SLE) patients, compared with healthy individuals. L1 hypomethylation and expression were correlated with SLE status. However, they found that the methylation levels of the LTRs of HERV-K and HERV- E did not influence on SLE disease status. However, it should be noted this experiment was purely done through computational analysis (Sukapan *et al.*, 2014).

Hypermethylation is also associated with induction of expression of HERV; Hu *et al.* (2017) demonstrated that ERV-9 was 90% methylated in the β globin gene in human erythroid progenitor cells EryP. ERV-9 methylation apparently regulated the downstream transcription of the β globin gene. In addition, the study showed that the knockout of the ERV-9 LTR, using CRISPR-cas9, reduced transcription downstream of this gene.

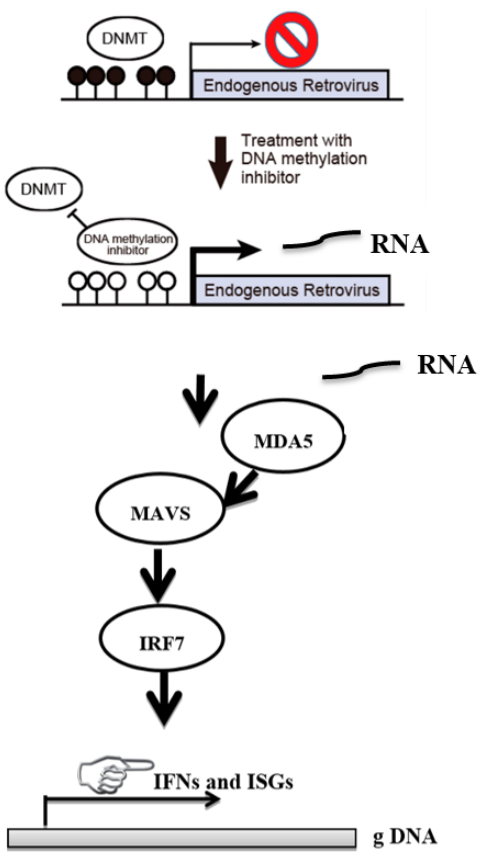


Figure 1.17: Activation of an anti-viral immune response in cell.

In a normal state, the 5′ LTR sequences of endogenous retrovirus (ERVs) are heavily methylated and HERV is silenced. However, the hypomethylation of 5′ LTR region induce dsRNAs that leads to the induction of an anti-viral immune responses such as the activation of interferon-response genes. Modified from (Nakaoka *et al.*, 2017).

In conclusion, the methylation state may regulate HERV-K LTR activity in melanoma and other cancer because transcription factors can play an important role in the expression of HERV-K. In addition, the use of inhibitors of methylation could help to find a good approach for the treatment of cancer. However, the relationship between HERV-K expression and epigenetic modification is not clearly understood.

1.19 Aims and Objectives of project:

1.19.1 Aims

The central aim of the project is to establish whether published connections between HERV-K expression and cellular phenotype are common in melanoma cell lines or specific to individual clonal lines. As other melanoma cell lines did show this connection we investigated HERV-K RNA and protein expression, and their relationship to the cellular phenotype and HERV-K epigenetic status.

1.19.2 Objectives

- HERV insertions upregulated in the melanoma cell lines will be assessed for utility as histological markers of HERV activation in melanoma.

- To explore the correlation between HERV-K expression and the suspension phenotype reported for some cell lines by (Serafino *et al*, 2009).

- To Investigate HERV-K mRNA and protein expression levels in melanoma cell, breast, colon and prostate cancer cells in different culture conditions.

- To explore the correlation between HERV 5' LTR expression and epigenetic state.

- To discover a novel HERV provirus polymorphisms (solo LTR/ provirus polymorphism / provirus presence / absence polymorphism) in genomic DNA using genome-wide HERV-specific amplification methods and NGS for sequencing.

- To compare the transcriptome of different melanoma cell lines and melanocytes in different growth conditions, using RNAseq.

2 Materials and Methods

2.1 Materials

2.1.1 Chemical reagents, laboratory equipment and consumables

Chemical reagents and laboratory equipment used in this project were supplied by one of the following suppliers: AB gene (Epsom, UK), Agilent Technology (USA), Ango (Austral), Ambion, (USA),Applied Biosystems (Warrington, UK), Bio-Rad (Hemel Hempstead, UK), Boehringer (Ingelheim, Germany), Cecil Instruments (Cambridge, UK), Clare Chemical research (Delores, USA), Clontech (Palo Alto, USA), Dako (Denmark), DigitalBio (Korea), Eppendorf Scientific (Hamburg, Germany), Fisher Scientific (Loughborough, UK), Fisons (Beverly, USA), FMC Bioproducts (Rockland, USA), Greiner Bio One(Glasgow, UK), Hybaid (Teddington, UK), Invitrogen (Paisley, UK), KAPA Biosystems (Woburn, USA), MJ Research (Waltham, USA), New England Biolabs (Hitchen, UK), Nagle Nunc International (Hereford, UK), New Brunswick Scientific Co. (New Jersey, USA), Oxoid Ltd (UK), Perkin Elmer (Cambridge, UK), Qiagen LTD (Crawley, UK), Serva, Sigma Aldrich (Pool, UK), Starlab (Milton Keynes, UK), Syngene Thermo Shandon (Pittsburgh, USA), USB (Staufen, Germany), UVP Life Sciences (Cambridge, UK) and Zymo research (Cambridge, UK).

2.1.2 Enzymes

The restriction enzymes *Ase*I and *Eco*RI were used for this project, and stored -20 °C. These enzymes were utilized depending on the reaction mixture, with the buffer provided by the manufacturer. Both were supplied by New England Biolabs. *Taq*I was supplied by Fermentas (York, UK). Taq and Pfu DNA polymerases were supplied by KAPA Biosystems (Woburn MA, USA).

2.1.3 Molecular weight markers

The markers used were 50 bp, 100 bp, ϕX174HaeIII and 1 kb Ladders. All markers were supplied by New England Biolabs.

2.1.4 Oligonucleotides

Oligonucleotides were synthesised by Sigma-Aldrich Company (Poole, UK).

Oligonucleotides (PCR primers) were designed using the HERV K107 and genome map and the human genome reference sequence (Appendix1). By Dr Richard Badge and Dr Catriona Macfarlane. The *pol* primer sequence was taken from Serafino *et al*. (2009). Primers were synthesised by Sigma-Aldrich Company (Poole, UK). They were diluted to a concentration of 50mM in TMT (10 mM Tris-HCl, pH 8.0) under PCR clean conditions (in a PCR hood). Primer stocks were stored at -20°C **(**Table 2.1).

Table 2. 1 The sequence of primers (5′- 3′) used in this project

Name	Sequence
ASE1	CGTTCAGCATATGGAGGATC
CMK107B	TAATTCTCCGCACACCACTG
CMK3LTR	GATCCTCCATATGCTGAACG
CMKENV	CAGGTGTACCCAACAGCTC
COBRAH	TTTATAGGTGTGTAGGGGTAATTTATTTT
COBRAR	GTAAAATTAAAAGTAAATATGTTTTTTATT
GAPDHF	GTCAAGGCTGAGAACGGGAA
GAPDHR	TCGCCCCACTTGATTTTGGA
IASPOLF	CCA CTG TAG AGC CTC CTA AAC CC
IASPOLR	GCT GGT ATA GTA AAG GCA AAT TTT TC
PGK1F	AAGTGAAGCTCGGAAAGCTTCTAT
PGK1R	AGGGAAAAGATGCTTCTGGG
RBD3	TACGTTCTATGCTAC
RBM13F	GTTTTCCCAGTCACGAC
RBM13R	CAGGAAACAGCTATGAC
RBMSL2	GTGGCGGCCAGTATTCGTAGGAGGGCGCGTAGCATAGAACG
RBX4	GTGGCGGCCAGTATTC
RBY1	GAGGGCGCGTAGCATAGAAC
SDHAF	TGGGAACAAGAGGGCATCTG
SDHAR	CCACCACTGCATCAAATTCATG
TERTC228-250 A	GGCCGATTCGACCTCTCT
TERTC228-250 B	CCACTACCGCGAGGTGCT

2.1.5 Chemicals, reagents, buffers and solutions

All routinely used chemicals, reagents, buffers and solutions were prepared as described in Table 2-2 and stored at room temperature, cold room 4°C or -20°C.

Table 2. 2 All chemicals, reagents and buffers recipes used in this project

Solutions	Preparation
10 X PCR Buffer	50 mM Tris HCl, 0.5 mM dNTPs, 12 mM NH_4SO_4, 125 µg/ml BSA, 6.7 mM β-mercaptoethanol, 2.5mM $MgCl_2$
10 X TBE	44.5 mM Tris-borate (pH8.3), 1mM EDTA.
11.1 X PCR Buffer	45 mM Tris-HCl, 11mM NH_4SO_4, 5mM MgCl2, 6.7 mM β-mercaptoethanol, 113ug / ml BSA, 0.05 mM EDTA, 1.1 mM dNTPs,
1x NEB3 buffer	50 mM Tris HCl pH 7.9, 10 mM MgCl2, 100 mM NaCl, and 1mM DTT
20x SSC	3 M NaCl, 0.3 M Tri-Sodium Citrate.
Ampicillin	100 mg/ml (dissolved in dH_2O).
Church Buffer	250 ml 14% SDS, 250 ml NaHPO4, 1 ml 0.5M EDTA (pH8.0). Filter Sterilised (0.4µm, Acrodisc)
DNA Loading Dye (6x)	4 ml 10xTBE, 3.5 g sucrose, 10 mg bromophenol blue. Final Volume to 10 ml with dH_2O
Ethidium Bromide	10 mg/ml (dissolved in dH_2O)
Fixative Solution	3:1 ratio of methanol: acetic acid. Make on day of use
IPTG	0.48 g of IPTG in 20ml of sterile distilled water.
Luria Bertani (solid and broth) media	Provided by Media Kitchen in the Department of Genetics, University of Leicester.
Single Molecule Dilution Diluent (SMDD)	5 mM Tris HCl pH7.5, 5 ng/ul sonicated *E.coli* genomic DNA
TMT (10mM Tris HCl pH 7.5)	TMT (10 mM Tris HCl pH 7.5)
Xgal	20ml of DMF, add 1g of Xgal (5-Bromo-4-Chloro-3-Indolyl β-D-galactopyranoside).

2.1.6 Cell lines and Materials for Tissue Culture

Table 2.3 describes the cell lines used in this work and their provenance. All equipment and materials required for cell culture experiments is described in Table 2.4. Table 2.5 describes the preparation of buffers used in the Tissue Culture section.

Table 2. 3 Details of cancer cell lines

Cell lines	Cell type	Type of cancer	Tissue origin	Type of media	provided by
A375P	Epithelial	Melanoma	54 years, female	DMEM	Dr Howard Pringle
A375M	Epithelial	Melanoma	54 years, female	DMEM	Dr Howard Pringle
SKMel5	Epithelial	Melanoma	24 years, female	RPMI-1640	Dr Eugene Tulchinsky
SKMel28	Epithelial	Melanoma	51years, male	RPMI-1640	Dr Eugene Tulchinsky
HEMaLP	Epithelial	Melanocytes	Light pigment from adult skin	245	Cascade Biologics®
SW480	Epithelial	Colon	50 years, male	DMEM	Dr Eugene Tulchinsky
LNCAP	Epithelial	Prostate	50 years, male	RPMI-1640	Dr Eugene Tulchinsky
T47D	Epithelial	Breast	54 years, femalr	RPMI-1640	Dr Eugene Tulchinsky
NTera2D	Epithelial	Embryonal carcinoma	22 years, male	DMEM	Genomic DNA provided by Dr Badge

Table 2. 4 Details of Materials and Equipment product numbers.

Equipment and Material	Product no.	Company
20% Sodium Dodecyl Sulphate solution (SDS)	BP1311-1	Fisher Scientific
30%acrylamide/bis acrylamide	EC-890	Proto Gel, UK
5-Aza-2dexocytidine 98%	2811	
96-well Plate	655180	Fisher Scientific
Ammonium Persulfate (APS)	A3678	Fisher Scientific
Bovine serum albumin (BSA)	9048-46-8	Fisher Scientific
Bromophenol blue	13-8026	Sigma-Aldrich
Cell viability plate	655983	Greiner Bio-One
Citric acid	1165418	Sigma-Aldrich,
CL-XPosure Film	34089	Fisher Scientific
Crystal violet 50 g	11K3694	Sigma-Aldrich,
4',6-diamidino-2-phenylindole (DAPI)	D9542	Thermo Scientific
Dimethyl sulfoxide (DMSO)	D5879	Sigma-Aldrich, UK
Disposable Neubauer haemocytometers (C-CHIPs	DHC-N01	Labtech International Ltd
Ethylenediaminetetraaceticacid (EDTA)	E5134	Sigma-Aldrich, UK
Fetal Bovine Serum(FBS)	EU-000F1	Seralab, UK
Glycerol	G5516	Sigma-Aldrich,
Glycine	G7126	Sigma-Aldrich,
Human Melanocyte Growth Supplement (HMGS)	S002-5	Fisher Scientific
Hydrochloric acid	J/4270117	Fisher Scientific
Immobilon-P polyvinylidene fluoride (PVDF) membrane	IPVH-00010	Fisher Scientific
MEK inhibitor UO126	V112A	Promega
Methanol	M/3950117	Fisher Scientific
N,N,N',N'-Tetramethylethylenediamine (TEMED)	T9281	Sigma-Aldrich,
Paraformaldehyde	30525-89-4	Fisher Scientific

Ponceau-S	D-3504	Sigma-Aldrich
Phosphate Buffered Saline(PBS)	BR0014G	Oxoid Ltd
Pre-stained protein ladder	26619	Fisher Scientific
Sodium Chloride Na Cl_2	57653	Sigma-Aldrich
Tris base	BP152-1	Fisher Scientific
Triton x100	106K0177	Sigma-Aldrich
Tween-20	P2287	Sigma-Aldrich,
Whatman Filter paper	3030917	Sigma-Aldrich
β-mercaptoethanol	M3148	Sigma-Aldrich,

Table 2. 5 Recipes of Buffer Preparing.

Buffer	Component
Fixation buffer	3:1 ratio of methanol: acetic acid. Make on day of use
Laemmli buffer (4x)	8% SDS, 40% (v/v) glycerol, 20 mM Tris pH 6.8 ,40 ml 20% SDS, 40 ml glycerol, 20 ml 1M Tris pH 6.8. Top up to 100 ml with dH_2O
Phosphate buffered saline (PBS)	g/L:- 8.0 sodium chloride; 0.2 potassium chloride; 1.15 di-sodium hydrogen phosphate; 0.2 potassium dihydrogen phosphate; pH 7.3± 0.2 at 25 °C. Dissolve 10 pBS tablets per 1 L of UltraPure water
Ponceau S Solution	0.1 % (w/v) Ponceau S in 5 % (v/v) acetic acid, 1g Ponceau S in 50ml acetic acid, made up to 1 L using distilled water
Protein Loading dye	0.2 % (w/v) bromophenol blue; 1% (v/v) β-mercaptoethanol
Protein Running Buffer (10x)	247.7 mM Tris, 1.9 M glycine, 1% (w/v) SDS ,30 g Tris base, 144 g Glycine, 50 ml 10% SDS top up to 1 L using dH_2O
Protein Running Buffer (1x)	24.7 mM Tris, 192 mM glycine, 0.1 % (w/v) SDS,1:10 dilution of 10x buffer
Protein Transfer Buffer (10x)	247.7 mM Tris, 1.92M glycine,30 g Tris base, 144 g glycine, top up to 1 L with dH_2O.
Protein Transfer Buffer (1x)	24.7 mM Tris, 192 mM glycine, 20 % (v/v) methanol
TBST/BSA (5%)	TBST, 5% (w/v) BSA
TBST/milk (5%)	TBST, 5% (w/v) dried skimmed milk powder
Tris Buffered Saline-Tween 20 (TBST, 1x)	5 mM Tris, 150 mM NaCl, 0.1% (v/v) Tween® 20,500 ml 20x TBS to 9 L dH_2O. Add 10 ml Tween ® 20. Top up to 10 L.
Tris-buffered saline (TBS, 20x)	1 M Tris, 3 M NaCl, Dissolve 121.14 g Tris and 175.32 g NaCl into 700 ml of dH_2O, then pH to 7.65 using concentrated HCl and top up to 1 L dH_2O.
trypsin EDTA 1x	0.5 mg/ml Trypsin, 0.22 mg/ml EDTA in PBS, pH 7.5

2.1.7 Primary and secondary antibodies

Table 2.6 describes the antibodies used in western blotting (WB), immunofluorescence (IF), immunohistochemistry (IHC) and their dilutions.

Table 2. 6 Primary and secondary antibodies.

Antibody	Product no.	Isotype	MW (kDa)	Dilution	Company
HERV- K *gag* (Primary)	1841-5	Mouse monoclonal IgG1	82	1:2000 WB 1:250 IHC 1:100 IF	AUSTRAL Biologicals
HERV- K *env* (Primary)	1811-5	Mouse monoclonal IgG2a	100	1:1000 WB 1:250 IHC 1:100 IF	AUSTRAL Biologicals
P_ERK (Primary)	4377	Rabbit monoclonal IgG	44,42	1:1000 WB	Cell signalling
α-tubulin (Secondary)	T5168	Mouse monoclonal IgG1	55	1:20000WB	Sigma-Aldrich
Anti-Mouse immunoglobulins / HRP (Secondary)	P 0447	Polyclonal goat	-	1:3000 WB	Dako
Anti-Rabbit immunoglobulins/HRP (Secondary)	P 0448	Polyclonal goat	-	1:2000 WB	Dako
Alexa Fluor® 488 Anti–Mouse IgG (Secondary)	A21202	Polyclonal donkey	-	1:500 IF	Invitrogen
Alexa Fluor® 594 Anti–Rabbit IgG (Secondary)	A21207	Polyclonal goat	-	1:500 IF	Invitrogen

2.1.8 Web services and software used for data analysis

- UCSC: The human genome browser at UCSC (Kent et al., 2002).
- Repeat masker: http://repeatmasker.org (A.F.A. Smit, R. Hubley & P. Green RepeatMasker).
- Primer3 on the WWW for general users and for biologist programmers. (Rozen and Skaletsky, 2000)
- MethPrimer: http://www.urogene.org/methprimer/index1.html (Li and Dahiya, 2002).
- Galaxy: http://main.g2.bx.psu.edu/ (Blankenberg et al., 2010).
- the Integrative Genomics Viewer (IGV) (Robinson *et al.*, 2011)

- Muscle http://www.ebi.ac.uk/Tools/msa/muscle/ (Edgar R.C., 2004).
- ImageJ software version 1.44 (Available from the Department of Genetics, University of Leicester), Image quant TL version 8.0.

2.2 Cell culture techniques

2.2.1 Cell culture

Cell culture work was carried out in a Class II laminar flow cabinet, with cell lines maintained in an incubator at 37 ºC, 5% CO_2 and 100% humidity. The range of cell lines used can be found in Table 2.3, including guidance on the type of cell culture media used. All media (DMEM and RPMI-1640), had 10% v/v FBS added to media. Medium 254 for melanocytes was supplemented with HMGS, as well as, 10% v/v FBS. Media was stored at 4 ºC and warmed to 37 ºC prior to use. Cells were maintained in culture and passaged at approximately 70% confluence.

Adherent cells were sub-cultured with three washes of phosphate-buffered saline (PBS), prior to trypsinization with 1 x trypsin EDTA. Cells were returned to the incubated at 37 ºC for 3 to 5 minutes (min), after which the flask was gently tapped to detach cells. The trypsin EDTA was neutralised with the addition of cell culture media. Cells were transferred to a 50 ml centrifuge tube and pelleted via centrifugation at 180 ×g for 5 minutes and the pellet re-suspended in 1 ml of media. The cells were seeded at the required cellular density in the correct sized flask and media added. Cells were immediately returned to the incubator.

Melanoma and other cancer cell lines were shifted from optimal growth factor conditions (10% FBS) to low-serum condition (1% FBS). The cells were seeded at a density of 2 x 10^5 cells in T 25 cm^2 flask and grown in 10% FBS. After 24 hours the

medium was changed to 1% FBS and cells were observed for detachment for 4-7days. After incubation, two phenotypes were shown, adherent and suspension. For sub-culture, the same process was done with 10% FBS and cells seeded at the correct density with the required culture media added (Table 2.7).

Table 2. 7 Volumes of PBS, TE and DMEM required for Cell Passaging.

Flask (cm^2)	PBS (ml)	Trypsin/EDTA (ml)	Complete DMEM added to neutralise TE (ml)	Total Volume of Media (ml)	Seeding density
T25	2	1	4	5	$2.5*10^5$
T75	5	2	5	10	$1.0*10^6$
T175	10	3	7	25	$3.0*10^6$

2.2.2 Freezing/Thawing cells

For maintenance in liquid nitrogen, cells were detached, pelleted and re-suspended in freezing media (80% FBS, 10% DMSO) at a concentration of approximately 2-5 x 10^6 ml cells and 1 ml volumes aliquoted into individual cryotubes. Cells were frozen at a rate of 1 °C/min in a -80 °C freezer and then placed in liquid nitrogen. For resuscitation, vials were thawed in a 37 °C water bath, centrifuged at 300 g and seeded at the appropriate density.

2.2.3 Cell counting

The harvested cells were re-suspended in complete media at a concentration of approximately 0.5-2 x 10^6 cells/ml. 10 µl of suspension was added to one chamber of a disposable Neubauer haemocytometer and the average number of cells calculated from 4 large squares (1 mm / square). This average was then multiplied by 10,000, to give a final concentration in cells/ml.

2.2.4 Chemical treatments

2.2.4.1 p-ERK treatment

A375P, A375M, SKMel28, SKMek5 and melanocytes cells were plated at approximately 5 x 10^5 cells per 10cm dish in 10ml of media. The cells were incubated for 48 hours and then media was aspirated, the cells were washed two times with PBS and p-ERK inhibitor (UO126) 10 µM and 20 µM was added to each dish. DMSO (the solvent for p-ERK inhibitor) was added to one dish as a control. The following day, lysis buffer was applied to each dish and lysates were collected for western blot analysis. Melanoma and melanocytes were grown in triplicates for each condition, as well as a control.

2.2.4.2 5-Aza-2dexocytidine (5AzdC) treatment

Melanoma and melanocyte cells were seeded at approximately 5 x 10^5 cells, the day before treatment into 10 cm dishes with 10% FBS and 1% FBS cultures. The media was then aspirated, and the cells were washed two times with PBS. 5AzdC was dissolved in DMSO to a final concentration of 10 µm /mL and stored in aliquots at −20°C. The 5AzdC treatment added to each dish at two concentrations, 2 and 10 µm, and DMSO alone was added to one dish as control. The cells were exposed to 5AzdC for 72 hours to allow the drug to be incorporated into DNA. Tissue culture medium was changed every day for both control and treated cells, to maintain drug stability during treatment. Melanoma and melanocytes were grown in triplicates for each condition, including a control. Lysates were collected for Western Blot analysis.

2.2.5 Protein analysis and Western Blotting

2.2.5.1 Analysis of protein

Lysates of cells were gently washed with PBS and lysed in Laemmli buffer, scraped into an eppendorf tube, and heated at 95 °C for 10 min. Lysates were sonicated 30 seconds (sec) to disrupt chromosomal DNA and centrifuged at 11,000 g for 1 min to pellet DNA material. Lysates were then stored at -20 °C or used for protein quantification.

2.2.5.2 Protein quantification

Bicinchoninic Acid (BCA) (Sigma) assay was performed according to the manufacturer's protocol. Quantification was carried out in a 96-well format. Briefly, 200 µl of BCA reagent was required for each sample, consisting of 50 parts BCA Reagent A to 1 part BCA Reagent B. Moreover, a series of protein standards was used, ranging from 2 µg/µl to 25 ng/µl of bovine serum albumin (BSA).

5 µl of pre-boiled protein sample/protein standard was added to the colorimetric reaction. A blank sample was also included, whereby the protein sample was replaced with Laemmli buffer.

The 96-well plate was placed on a shaker for 30 seconds (sec), then at 37 ºC for 30 min. The plate was submitted to absorbance reading at 562 nm, taken on a BioTek ELx808 Absorbance Microplate Reader. The absorbance value obtained for the blank sample was subtracted from the absorbance readings for the protein samples and standards. A standard curve was generated for the protein standards, allowing the concentration of the unknown protein samples to be determined.

2.2.5.3 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to determine the expression of a target protein. The percentage of the resolving gel determined according to the molecular weight of the target protein, the ranges used were from 6 to 15% (Harlow and Lane, 1988). In our experiments, the percentage of resolving gel was 8% and 10 ml final volume was made by mixing according to Table 2.7. The final mixture of solutions was transferred to the pre-assembled gel apparatus and isobutanol was added (400-500 µl) to the gel to create a homogenous surface. After setting up the gel, the isobutanol was removed and the gel surface washed 3 times with water and dried by tissue carefully. The next step was 5ml of stacking gel (5%) was made by mixing (3.4 ml water, 0.8 ml acrylamide mix, 1.25 ml 1M Tris-HCl (pH 6.8), 0.01ml 10% SDS, 0.01 ml Ammonium Persulfate (APS) and 0.01 ml N,N,N',N'-Tetramethylethylenediamine (TEMED) and poured on top of the resolving gel. Directly, a 15 well comb was put into the stacking gel before it became solid. A BioRad running tank was used to electrophorese the gel, filled with 1x running buffer. At the same time, samples were heated at 95°C for 5min. The comb was removed from the stacking gel and 30µg protein samples were loaded carefully onto each well. The protein marker was loaded into the first and last wells of the gel. After that, the samples were run at 140 volts for 80min.

Table 2. 8 The preparation of solutions to make 8% resolving gels.

Gel (%)	Water	30% acrylamide	1.5 M Tris (pH 8.8)	10% APS	TEMED
8	4.6	2.7	2.5	0.1	0.006

2.2.5.4 Western blotting process

After SDS-PAGE was completed, the resolved gel was rinsed in water, the proteins transferred to a polyvinylidene difluoride (PVDF) membrane using a semi-dry transfer apparatus. For this six pieces of 3 mm Whatmann filter paper and one piece of 0.2 µm pore PVDF were cut to match the size of the gel and a sandwich created. In a tray, three pieces of filter paper were wet in 1 x transfer buffer and placed in the centre of the transfer apparatus followed by the piece of PVDF, previously activated by immersion in in methanol. Then the SDS-PAGE gel was put on top membrane and the final three pieces of filter paper wet by soaking in transfer buffer. Bubbles were removed by gently rolling a 5 ml pipette over the sandwiched gel and membrane. The transfer cassette was properly orientated within the transfer tank, which was filled with transfer buffer. Protein transfer occurred at 25 volts, overnight (Figure2.1). After transfer, Ponceau S staining was used to check successful protein transfer and then the membrane was washed with water to remove the staining.

Membranes were then incubated with 1 x blotting buffer (5% w/v BSA powder in TBS-T), rotated for 55 min, washed three times with TBS-T (each wash was 5 min long) and incubated with the primary antibody diluted in blotting buffer, overnight in a cold room. After that, the membrane was washed three times in TBS-T, and then incubated for 45 min in secondary antibody diluted in blotting buffer. It was then washed three times with TBS-T before application of the enhanced chemiluminescent reagent (ECL from thermo scientific). To do this excess solution was drained and each membrane placed protein-side down on the ECL reagent (1ml /5 cm2 membrane) for 1 minute. ECL reagent was drained off, membranes placed within a plastic film, and then positioned in an autoradiography cassette. X-ray film was placed over the membrane for a variety of exposure times and developed using an AGFA Curix 60 film developer. Films were

scanned with a Canon LIDE 60 scanner. Samples were processed as three biological repeats with triplicate technical replicates.

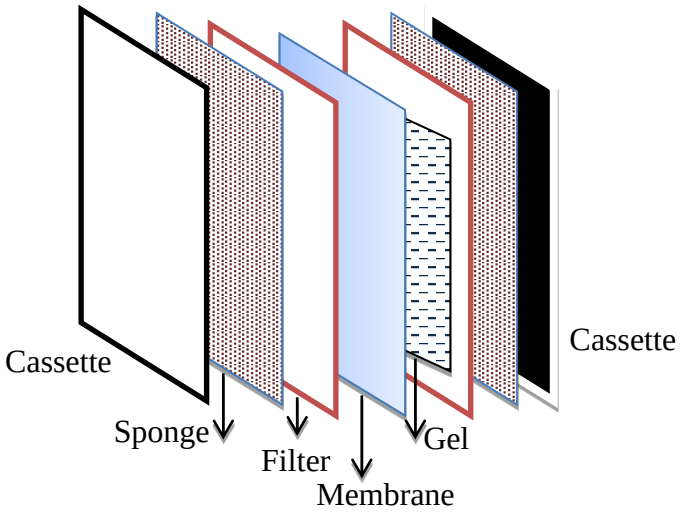


Figure 2.1: protein transfer set up.

The schematic represents the set-up of a wet transfer of proteins from a resolving gel, processed via SDS-PAGE, onto a PVDF membrane

2.2.6 Immunofluorescence

Immunofluorescence assays were used because this technique is useful to localize target proteins (*gag* and *env)* expression within cells. Briefly, melanoma cells (A375P, A375M, SKMel28 and SKMel5) and melanocytes cell (HEMaLP) were grown in two conditions 10% FBS and 1% FBS cultures.

100,000 cells were seeded on a cover slip (18×18 mm; thickness no.1; borosilicate glass), placed in a 6 well plate and incubated for 48 h before staining. After aspiration of media, cells were rinsed two times in PBS, the washing steps occurred on a shaker for 5 min. Then, 1ml 4% paraformaldhyde was added to fix the cells during incubation for 15 min at room temperature. After fixation, the washing step was repeated again. 0.5% Triton (1ml) was added for 5 min to permeabilize the cell membranes. The

washing step was repeated again and the coverslips incubated with 100µl primary antibody diluted in 3% BSA/PBS for 1h at room temperature in the dark (covered by aluminium foil). After washing, 100µl fluorescent-conjugated secondary antibody was added and incubated for 1h at room temperature. The cells were rinsed three times in BPS. To resolve nuclear staining, 1ml of 0.5µg/ml 4', 6-diamidino-2-phenylindole (DAPI) was applied, and incubated 5 min at room temperatures. After DAPI staining, the cover slips were washed twice and mounted into glass slide which included a drop of Fluoromount G (Thermo fisher scientific). The cover slips were then reverse mounted onto Menzel-Gläser microscope slides (25×75×1 mm). The slides were left to dry at room temperature and the edge of the cover slip was sealed with nail polish to prevent drying and movement under the microscope. Immunofluorescence microscopy was used to visualise antibody staining, and the images by were analysed with ImageJ. All samples were analysed in triplicate.

2.2.7 Immunohistochemistry

Immunohistochemistry (IHC) is an important method for localization of antigens / protein expression in tissue sections, by the use of specific antibodies. The validity and quality of IHC is impacted by the antibody quality, antibody dilution factor, and how the tissue is fixed and processed.

Two stages of the melanoma disease process were used, naevus and aggressive melanoma. The slides were provided by the Histopathlogy group in the Department of Cancer Studies, University of Leicester. In this experiment, IHC was carried out using the following protocol with the Novolink™ Polymer Detection kit. Additionally, primary antibody dilution was determined by using a series of dilutions of antibodies (*gag*, *env*) to find the best dilution factor for working with tissue sections. Dilution of antibodies is illustrated in Table 2.6.

The experiment included normal tonsil tissue as a positive control and a negative control which did not include primary antibodies, to exclude non-specific background staining. Briefly, the slides were incubated at 60°C for 30 min in an oven and cooled to room temperature. To remove paraffin from the tissue surface, the slide sections were incubated in xylene for 2 x 3 min, then rehydrated in decreasing concentrations of ethanol, absolute (2 x 3 min), and 95%, ethanol (2 x 3 min) and finally washed in water for 5 min. The next step, antigen retrieval was performed by heat treatment in 1 x Citric acid buffer solution pH 6. All slides were put into a microwave vessel which contained 500ml of citric acid buffer solution, and exposed in a microwave oven at full power (Power rating of microwave 100W) for 20 min. Following treatment, slides were left to cool in the solution at room temperature for at least 30 min and then the slides were transferred into a humidified chamber to neutralize endogenous peroxidase activity by adding 100ul of peroxidase block to all sections, for 5 min. The slides washed two times in 1xTBS-T, for 5 min. Protein block was added to slides in the same way for 5 min to block non-specific protein binding, followed by a washing step in 1xTBS-T. Dilutions of *gag* and *env* antibodies in 3% BSA were applied and slides incubated over night at 4ºC. The next day, the slides were washed twice with 1xTBS-T for 5min, then a post-primary block was applied, 100µl for 30min at room temperature.

The washing step repeated again with TBS-T and 100µl of the polymer buffer was added for 30 min at room temperatures, followed by washing once with TBS-T for 5 min. Appropriate amounts of 3, 3 – diaminobenzidine (DAB) substrate solution (100µl), made freshly by mixing 1 volume of chromagen with 20 volumes of substrate (1:20), was added for 5 min and directly washed with tap water for 5min. For staining, Mayer's haematoxylin was applied for 30 sec followed by washing with tap water. The first step in this process (hydration) was then repeated, but in the opposite direction to dehydrate

the slide. This was started with 95% ethanol (1 x 5min), then 99 % ethanol (2 x 2min), and then xylene (2 x 3min). After drying the slides, coverslips were mounted over tissue sections by using DPX (DPX is clear and colourless reagent and help to view sections under a microscope). The slides were kept to dry for 30 min at room temperatures before analysis by microscopy.

2.2.8 Viability assay

The RealTime-Glo™ MT Cell Viability Assay was used to ensure that cells in 1% FBS showing a suspension phenotype were viable cells. A375P, A375M and SKMel5 were used in these experiments. Adherent cells under 1%FBS were compared to suspension cells and at the same time cells in 10%FBS were used as control.

The principle of this method is the NanoLuc® Enzyme that produces a luminescent signal depending on the number of living cells (Figure 2.2). This is an assay that can be used to determine if observed cells are still alive and growing under starvation, or simply dead and dying cells remnants.

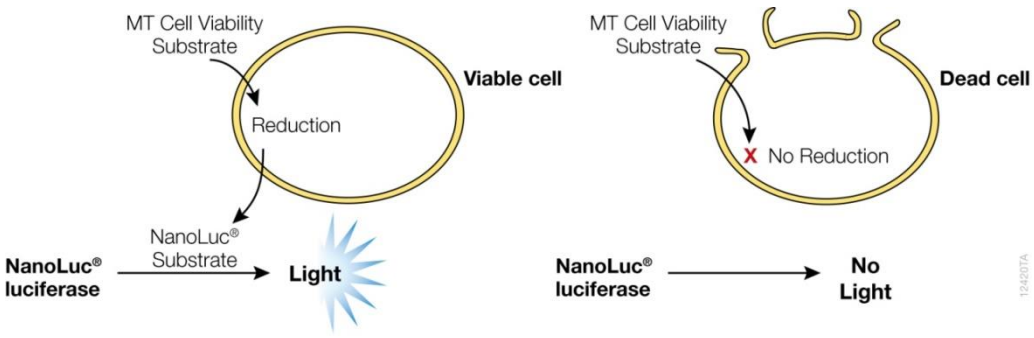


Figure 2.2: RealTime-Glo™ MT Cell Viability Assay principle.

NanoLuc® luciferase and MT Cell Viability Substrate were added to cells in culture. The MT cell viability Substrate is reduced to a NanoLuc® substrate by the metabolic activity of living cells. The final substrate reactant diffuses from living cells into the medium and is directly used by the NanoLuc® Enzyme to produce a luminescent signal. (Figure from the supplied Promega protocol (protocol reference G9711).

This experiment was done according to the manufacturer's protocol. Briefly, cells (100 µl / well) were seeded at s $1x10^3$ cells/ml into 96 well white walled tissue culture plates with a clear bottom, and allowed to adhere for 24h, until approximately 30% confluent. The next day, the MT Cell Viability Substrate and NanoLuc® Enzyme were warmed to 37°C in a water bath, and the MT Cell Viability Substrate and NanoLuc® Enzyme added to the medium at a final concentration of 2 x for each reagent. Then the media was aspirated from each well and replaced with media which included the cell viability assay reagents. Following a 1 h incubation period with the RealTime-Glo™ MT Cell Viability reagents, the 96 well plate was measured by a plate-reading luminometer, then returned to the incubator.

Measurement were taken at different time points -24 h 48 h and 72 h. This experiment was done with 4 repeats for each cell type / condition and with three biological replicates and with blank control wells containing media only.

2.2.9 Colony formation assay

The principle of this method is that individual cells were plated in a dish and allowed to form a colony of cells that is the stained with crystal violet (0.5% w/v). The number of colonies is related to the cell proliferation level. Colony formation was used to compare the proliferation capability between adherent cells in culture under 10% FBS culture and 1% FBS culture as this assay is not suitable for cell showing a suspension phenotype. Briefly, cells were seeded in 6 cell plates at different density: 1000 cells / well in 10% FBS, and 2,000 cells / well in 1% FBS. The cells were incubated for 7 days at 37°C. Then the media was aspirated off, the wells were washed 2 times with PBS, and the colonies fixed with fixation solution (Acetic acid/methanol 1:3 (vol/vol)) for 15

min. Fixation solution was discarded, and 1 ml of a solution of crystal violet (0.5% w/v) was added to each well for 20 min at room temperature. Stained cells were washed with tap water until a clear background was visible. Plates were air-dried. Finally, the number of colonies in each well was calculated and cell morphology was photographed, to quantify the colony formation ability. Colony formation was calculated using Plating Efficiency (PE) according to Franken *et al.* (2006), which calculated as follows: PE = no. of colonies formed / no. of cells seeded x 100%.

2.3 Molecular biology techniques

2.3.1 DNA extraction

DNA was extracted from the cell pellets at approximately 3-5 x 10^6 cells, and used the blood and tissue DNA extraction kit on the Maxwell® 16 Instrument, applying the buffy coat protocol. Briefly, thawed cell pellets were carefully transferred using a wide bore pipette tip to chamber 1 (Lysis) of the blood and tissue DNA extraction cartridge. The cartridge was then processed by the Maxwell 16 Instrument. The purified DNA was eluted in ~300 µl of elution buffer. DNA quantity and purity were assessed by UV spectrophotometry.

2.3.2 PCR technique

All PCR setup was done in a class II laminar flow hood to ensure that all the reagents were kept PCR-clean. The reactions of PCR mix were made up to a final volume of 10 or 20 µl per reaction. Each reaction included 0.02 U/µl of Taq DNA polymerase, 1× PCR buffer (11.1 × or 10 x), 0.5 µM of each primer, and water was added to bring final volume to 10 or 20 µl. High molecular weight genomic DNA (20ng) that had been minimally freeze-thawed was used for all PCRs. Specific primers names and conditions

are associated with each assay, but an example PCR reaction is tabulated below. These primers were used to amplify regions of the HERV-K *pol* gene. The PCR product was expected to be approximately 400bp in length. The primer combinations and cycle conditions are shown in Table 2.9. These conditions were used, unless otherwise specified.

Table 2. 9 Shows the amplification HERV-K (*pol*) by PCR.

Amplification	5' Primer (Forward)	3' Primer (Reverse)	PCR conditions
Pol	IASPOLF	IASPOLR	96°C 1 min (96°C 30 s,62 °C 30 s,72°C 1 min) x34 cycle, 72°C 10 min,15 °C forever

2.3.3 Gel Electrophoresis

To confirm PCR reactions were successful and had produced a good yield of DNA or cDNA, PCR product was mixed with 5 x Loading dye, and run out on 1-2% LE agarose gels containing 0.5x TBE buffer and including 0.5ug/ml of ethidium bromide. The molecular weight markers were used depending on the size of the product in the running gel. PCR products were visualized under UV transillumination.

2.4 Sequencing

2.4.1 Sequencing reactions

10 µl final volume sequencing reactions were set up. These reactions comprised 1 µl Big Dye Terminator Ready Reaction (version 3.1), 1.5µl of 5x Big dye terminator buffer, 1 µl of 3.2 µM sequencing primer and 20-30 ng/kb of DNA or cDNA. The PCR conditions were (96^0C for 10s, 50^0C for 5s, 60^0C for 4min) x25. If the concentration of

DNA was low (under 2.5 ng/ul), 5 µl of DNA was added directly to the sequencing reaction.

2.4.2 Sequencing reaction clean-up

10 µl of water and 2 µl of 2.2% SDS (Sodium dodecyl sulphate) were added to each sequencing reaction. They were incubated in a PCR machine at 98^0C for 5min and 25^0C for 10 min. Then Edge Biosystems Performa DTR gel filtration columns were used according to the manufacturer's protocol to remove unincorporated dyes and reaction components. The eluted solution was labelled and sent for data collection by PNACL.

2.5 Southern blot

Southern blotting was used for determining the presence of specific DNA sequences in suppression PCRs. Gels to be blotted were treated with denaturing solution for 20 min with shaking. After that the denaturation solution was removed and the gel washed with distilled water. Then neutralizing solution was added and incubated for 20 min with shaking. The gel was put on a Southern blotting apparatus and the transfer reservoir filled with 20xSSC. Blotting paper and nylon membrane was cut to a slightly larger size as the gel. The nylon membrane was wet with 3xSSC and put on top of the gel. Two pieces of 3M paper were cut and soaked in 3x SSC. Paper towels covered the 3M paper and a glass plate was then put on the top and left overnight. After transfer, a pencil was used to mark the wells on the membrane, and the orientation mark. The membrane was dried at 80^0C for 10 min and then exposed to a UV cross linker (Stratagene) to fix the DNA.

2.6 Hybridisation

2.6.1 Random priming oligonucleotide radiolabelling

20 ng of purified 5' HERV K107 LTR DNA probe was labelled for each hybridization. Water was added to the probe DNA to bring the volume to 30 µl.

This dilution was boiled for 6 min and then chilled for 1 min on ice. 6µl of Oligo Labeling Buffer (OLB), 1.2ul of BSA (DNAse free at ~10mgml-1), 1ul of Klenow (5,000ml-1, USB) and 1.5ul of 32Pα d-CTP were added to the denatured DNA in the radioactive area, and incubated at 37^0C for 2-5 hours in a water bath.

The reaction was chilled on ice and, 70µl Oligo stop solution (OSS), 30µl of herring sperm DNA (3 mg ml-1), 30 µl 3M sodium acetate (pH 5.6) and 425µl of 100% ethanol were added to the reaction and mixed. After centrifugation for 3 min at 16,000 X g, 550 µl of 80% ethanol was washed over the pellet in the tube, and it was centrifuged for 1 min at the same speed as previously. The supernatant was removed and the pellet allowed to air dry. 600µl of water was added to the tube and boiled for 6min in a heating block. The denatured labelled probe DNA was chilled on ice for 1 minute then added to10 ml hybridisation solution. After dilution, the hybridisation solution containing the probe was added to the hybridisation bottle containing the cross-linked Southern blot.

2.6.2 Pre- hybridization and hybridization

The nylon membrane from the Southern transfer was placed on piece of nylon gauze and rolled up tightly. The membrane and gauze were placed in a hybridisation bottle in a hybridisation oven. The membrane was pre-hybridised with 50 ml modified

CHURCH solution hybridisation buffer at 65°C for 1 hour. The Church buffer was refreshed and the radiolabelled probe added to 10 ml hybridisation solution. The nylon membrane was incubated with the probe at 65 °C overnight, with constant rotation.

2.6.3 Post Hybridisation

After the hybridization was performed overnight at 65°C, the membrane was washed in three washing steps. The first wash was 30-50 ml 0.2x SSC and 0.5% SDS (500 ml = 5 ml 20 x SSC, 25 ml 10% SDS, 470 ml grey line H_20) and the membrane was washed for 15 min at 65°C. This step was repeated a second time. In the third washing step the hybridisation bottle was filled with 30-50 ml 0.1 x SSC and 0.1% SDS and washed for 30 min at 65°C. The membrane was then rinsed in 3 x SSC and allowed to briefly air dry, wrapped in Saran wrap. And put into a cassette with a phosphoimager screen and exposed overnight at room temperature. The pattern of radioactive signals was collected by the phosphoimager screen and scanned with the Typhoon 9400 (GE Healthcare, Life Sciences) or exposed to X ray film.

2.7 DNA extraction from agarose gels

To obtain DNA fragments from agarose electrophoresis gels, the DNA was cut from the gel in an agarose block. This extraction was done using the Zymoclean™ Gel DNA Recovery kit according to the manufacturer´s protocol. Briefly, after cutting the DNA band containing block on the Dark Reader visible light transilluminator (Clare Biosciences), to minimize UV damage, the gel slice was weighed and three volumes of ADB was added. The gel slice was incubated at 55°C for 10 min with gentle mixing to sure that all visible gel was dissolved. The gel solution was added to a collection tube, and centrifuged at maximum speed for 1 min. After discarding the flow through liquid,

200 µl of DNA wash buffer was added to column and it was centrifuged again. The flow-through liquid was discarded, and this washing step was repeated twice. The DNA was eluted by adding 6µl of DNA elution buffer directly to the column and centrifuging the column at 16,000 X g for 1 min. The eluted DNA was stored at -20°C.

2.8 Cloning PCR products

2.8.1 Ligation

The pGEM-T Easy vector system (Promega) was used for cloning DNA fragments. 10 µl final volume Ligation reactions were setup with 5 µl of 2x Rapid ligation buffer, 0.5 µl (25 ng) of pGEM-T Easy vector, 1µl of T4 DNA ligase and up to 3.5µl of PCR product. All ligation solutions were contained in the pGEM-T Easy PCR product cloning kit (Promega). A positive control reaction contained 2 µl control insert DNA supplied with the kit and a background control reaction included H_2O in place of the PCR product. The reactions were incubated overnight at 4^0C in the cold room.

2.8.2 Transformation

5 µl of the ligation reaction was added to 200 µl of *E.coli* competent cells (stored at -80^0C) and was thawed on ice, for 30 min mixing gently each 15 min. Then, the reaction was placed into a heat block at 42^0C for 30 sec. Directly, the competent cells were put on ice for 2 more min. 900 µl of pre-warmed (37^0C) SOC media + glucose (0.04 g / 10 ml) was added and this reaction was incubated at 37^0C for 45 min. After that 40 µl of Xgal (20ng/µl) and 20 µl of IPTG (0.15M) were spread on warm Ampicillin (200 µg / ml) containing LB agar plates. 200 µl of the transformation reaction were then spread on each plate, and once dried the plates were incubated overnight at 37^0C.

2.8.3 Colony PCR reactions for detecting plasmid inserts

Growing colonies of bacteria (1-2 mm) on Ampicillin containing solid media were picked with a sterile yellow tip and mixed with 100 µl of pure water.

The bacterial suspension was incubated at 98^0C for 5 min and 25^0C for 10 min, in a PCR machine to lyse the bacteria. Then, 10 µl PCR reactions were setup with 1X PCR buffer B, 0. 5 µM RBM13F and RBM13R primers, 0.02 U/µl of Taq DNA polymerase and 1 µl of lysed bacterial suspension. The master mixture was made in the hood and the bacterial lysate was added on the bench. This experiment included hood negative control and bench negative controls to ensure amplicons were specific to the added bacterial lysate. PCR conditions were 96^0C for 1min, (96^0C for 30 sec, 55^0C for 30 sec, and 68^0C for 1.5 min) x 30 and 68^0C for 10 min. The PCR product was run on a 2% gel to confirm amplification of the cloned insert.

2.8.4 Miniprep of plasmid DNA

For the extraction of plasmid DNA from bacterial clones, the Plasmid DNA mini kit II (from Omega) protocol was followed. Briefly, single colonies were incubated in 3ml LB-broth medium at 37°C, with shaking at 200rpm, overnight. Purification of DNA was done following the manufacturer's manual and the DNA was eluted with 50µlof elution buffer. DNA concentration was measured on the NanoDrop UV Spectrophotometer and the samples were stored at -20°C prior to sequencing.

2.9 RNA extraction

To extract RNA, cell pellets of approximately $5x10^6$ cells was used, and th Maxwell® 16 LEV simply RNA Cells Kit and Maxwell® 16 LEV simply RNA Tissue Kit was used for extraction RNA from melanoma and other cancer cells. The procedure was performed according to the manufacturer's protocol. Briefly, firstly a homogenization /

1-thioglycerol working solution was made by adding 20µl of 1-thioglycerol per ml of homogenization solution, and mixed well by vortexing. 200µl of homogenization /1-thioglycerol solution was made for each sample.

For preparing DNase I, 275 µl of nuclease-free water was applied to the vial of lyophilized DNase I, inverted 2-3 times until the solution was clear. Then, 5µl of Blue Dye added to the reconstituted DNase I, and mixed gently.

After preparing all working solutions, cartridge preparation was done. 200µl of homogenization /1-thioglycerol solution was added to each sample and mixed by pipetting, and then 200 µl of lysis buffer was added. Vigorous vortexing for 15 sec was carried out to mix the solution. All 400µl of the lysate was transferred to well one of the Maxwell cartridge, and 5 ul of DNase I was pipetted into well four. Also, a LEV Plunger was placed in each well number eight. Then, 50µl nuclease-free water was added to the bottom of each Elution Tube.

2.9.1 Quantification of RNA and DNA

Agilent RNA 6000 nano, Agilent DNA 1000 nano and DNA high sensitivity kits were used to for quantification of RNA and DNA. The Agilent 2100 bioanalyser instrument was used following the instrument manufacturer's protocol.

2.9.2 Treatment of RNA

The TURBO DNase Treatment and Removal Reagent System kit (from Invitrogen) was used to remove contaminating gDNA from RNA samples followed the manufacturer's protocol. Briefly, 10 µg total RNA from each sample was treated in a 50µl reaction. This reaction contained 10x TURBO DNase buffer, 1µl TURBO DNase, and was mixed gently. This reaction was incubated at 37ºC for 30 min. 5µl of resuspended DNase

inactivation reagent was added, and the reaction incubated for 5 min at room temperature. Then, the reaction was centrifuged for 1.5 min at 10,000 x g, and the supernatant transferred to a clean eppendorf tube.

2.9.3 Synthesis of cDNA and Reverse Transcriptase (RT) assay

After RNA treatment, the SuperScript® VILO™ cDNA Synthesis Kit was used to generate first strand cDNA, according to the manufacturer's protocol. Briefly, a master mix consisting of 4 µl of SuperScript® VILO™ master mix buffer, was combined with treated RNA (2 µg) and DEPC-treated water to bring the final volume to a 20 µl reaction. The reaction mixture was mixed gently and the tubes put in a PCR machine. cDNA synthesis conditions were 25°C for 10 min, 42°C for 60 min and 85°C for 5 min. Newly synthesized cDNA was then stored at –20°C until used for endpoint PCR and quantitative polymerase chain reaction (qPCR). In all cases a negative control was included, following the same steps except denatured reverse transcriptase enzyme was included by incubating the SuperScript® VILO™ master mix buffer and water at 65 °C for 10 min before adding the RNA sample.

Negative reverse transcriptase and negative RT assays (water instead of cDNA) were performed with each experiment. Additionally, a positive control was performed to assess the efficiency of the reaction.

2.9.4 Determination of efficiency of primers for qPCR

This protocol was provided by the Department of Cancer Studies, University of Leicester. In this experiment, cDNA from T47D (a breast cancer cell line) was used for the quantification of primers used in this project. Generally, primers for three housekeeping genes, Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), phosphoglycerate kinase 1 (PGK1) and succinate dehydrogenase complex flavoprotein

subunit A (SDHA) were used as the references genes and compared to the HERV *pol* gene primers. The standard curve consisted of a dilution series of factors of 1:2, with 8 points. The range covered started from the stock concentration (120ng/ul) solution and then 20 ng to 0.3125 ng. The primers (0.3µm), both the forward and reverse, were and then 5µl of Fast SYBR® Green Master Mix (2x) was mixed with each primer for each dilution and sterile water was added to make the final volume 10 µl. RT negative and PCR negative controls were included in this experiment. All samples were tested in triplicate, on 96-well PCR plates. The melt curve of primers during cDNA amplification showed good specificity, as illustrated in Figure 2.3. The efficiency of primers for each gene was plotted against the amount of cDNA input and a standard curve for each gene was calculated. According to the Applied Biosystems StepOnePlus Real-Time PCR Systems software, the slope of this curve (m) was used to calculate the efficiency (E) of the primers as follows formula: Efficiency (E) = 10^ (-1/slope). A good reaction should have an efficiency between 1.8 and 2.2, which corresponds to a slope between -3.58 and -3.10.

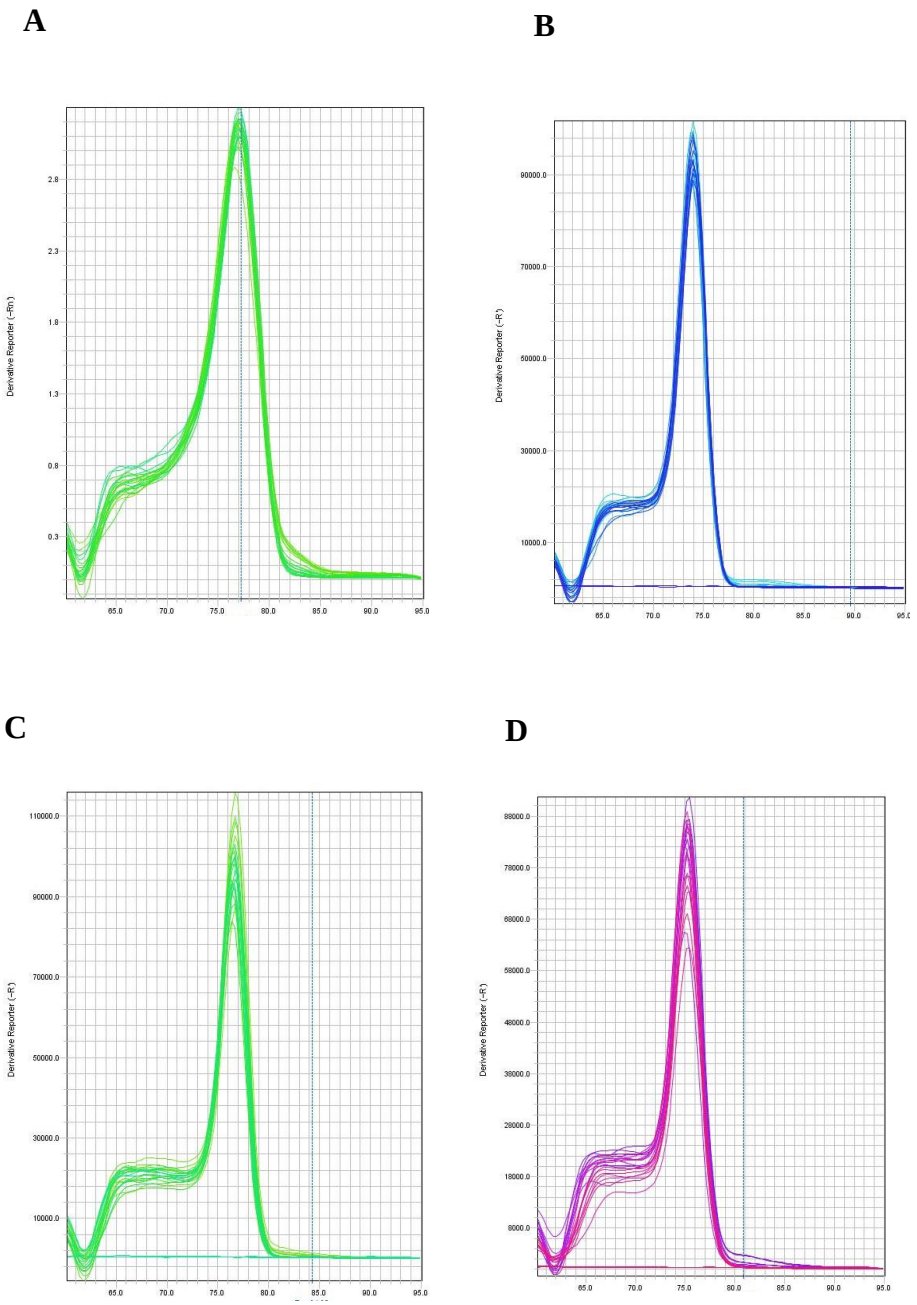


Figure2.3: Melt curve analysis for four targets.

These experiments were carried out on an Applied Biosystems StepOnePlus Real-Time PCR Systems instrument for 8 different points**.** The melt curve showed a single peak at a specific melting temperature for all primers, and a clear peak of fluorescence was observed repeatedly at the same melting temperature indicating a single PCR product. **(A)** *Pol* primer melt curve, efficiency (E) = 2.00. **(B)** SDHA primer melt curve, e efficiency (E) = 2.00. (C) GAPDH primer melt curve, efficiency (E) =1.97. **(D)** PGK1 primer melt curve, efficiency (E) = 2.2.

2.10 Quantification *pol* expression by qPCR

Amplification of specific PCR product (HERV-K *pol* mRNA) using the Fast SYBR® Green Master Mix protocol was carried out in triplicate for each sample, and three biological repeats for each grouped of cells grown in two conditions (10% FBS and 1% FBS), in 20µl reactions. This reaction included 1/10 dilution of cDNA, 0.3µm of the forward and reverse primer and Fast SYBR® Green Master Mix buffer (2X) 10 µl. The experiment was done with GAPDH, SDHA and PGK1 housekeeping genes were used as references. Within each experiment, no cDNA was used as negative control and a negative RT control was used to verify any contamination of gDNA. qPCR conditions as shown in Table 2.10.

Table 2. 10 qPCR conditions used to amplify HERV-K *pol* mRNA.

Steps	Temperature	Time (sec)	Cycle
Activation	95	20	40 X (Denature and Extend)
Denature	95	3	
Extend	60	30	

2.10.1 Evaluation of qPCR results

For this project, data obtained from qPCR were analysed to quantify gene expression levels for *pol* mRNA by comparing the cycle threshold (CT) values of the target genes in different culture conditions and for cell lines showing different phenotype with the reference genes. These values were normalised to the three reference genes. Under optimal conditions, amplification is exponential corresponding to a doubling of the amplified sequence during each cycle. Because this optimum is not always achieved,

the efficiency method of PfaffI was used (PfaffI, 2001). The CT value is defined as the number of cycles in the exponential phase of amplification as follows:

$$\frac{E^{(\text{Target control})} / E^{(\text{Reference gene control})}}{E^{(\text{Target treatment})} / E^{(\text{Reference gene treatment})}}$$

E = Efficiency of reverse transcription

2.11 DNA converted methylation method

2.11.1 Bisulphite sodium conversion treatment

The principle of this method is the denaturation of DNA and treatment with sodium bisulphite reaction. In this step unmethylated cytosines are converted to uracil while methylated cytosines remain unchanged. The subsequent PCR replaces uracil with thymine in the region of interest, directed by region specific primers. Sequencing of the PCR products reveals whether cytosines in the CpG dinucleotides were methylated in the original genomic DNA, as illustrated in Figure2.4.

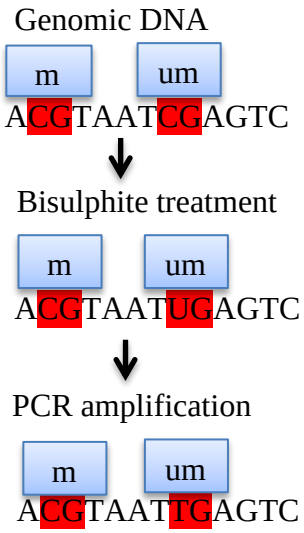


Figure2.4: Schematic representation of bisulphite conversion analysis of DNA methylation.

The EpiTect Bisulphite Kit from QIAGEN was used for DNA conversion, according to the manufacturer´s instructions. Briefly, gDNA (1µg) was used and the required numbers of bisulphite-mixes prepared by adding RNase-free water (800 µl) to each aliquot. The reaction mix of gDNA, with the solutions of from the Epitect kit are shown in Table 2.11. After mixing all reactions in the PCR tube, thermal cycling treatment was carried out as shown in Table 2.12.

Table 2. 11 Epitect Bisulphite conversion reactions.

Materials	Volume per reaction (µl)
DNA 1µg	Maximum 20
RNase-free water	Variable
Bisulfite Mix	85
Protect Buffer	35
Total	140

Table 2. 12 Thermal cycler conditions for bisulphite conversion.

Steps	Time(min)	Temperature
60°C	5	95°C
Incubation	25	95°C
Denaturation	5	95°C
Incubation	85	60°C
Denaturation	5	95°C
Incubation	175	60°C
Hold	Indefinite	20°C

After completion of the bisulphite treatment, the reaction was transferred to a new 1.5ml tube and 560µl buffer BL, containing 10µg/ml carrier RNA, were added. The mixture was then transferred to an EpiTect Spin column. The column was centrifuged for 1min at >14,000g. After discarding the flow-through, the column was washed by adding 500µl BW buffer and centrifuged for 1min at >14,000g. The flow-through was discarded. Then, the converted DNA was desulphonated by adding 500µl buffer BD and carrying out a 15min incubation at room temperature, with column lid closed. The column was then washed twice with 500µl BW buffer, transferred to a new 2ml tube and again centrifuged for 5min at >14,000g to get rid of any residual wash buffer. The column was transferred to a new 1.5ml tube and the converted DNA was eluted by adding 20µl EB buffer. Eppendorf tubes containing bisulphite-treated DNA were stored at (-20°C) until further processing.

2.11.2 COBRA assay

The Combined Bisulfite Restriction Analysis (COBRA) amplification was performed on sodium bisulphite converted DNA. 10 µl PCR reactions was setup, including 11.1 x buffer, specific primers for the methylated region (0.5µm) were used to amplify 222bp from HERV LTR region. Clean water was added to bring the final volume to 10 µl. PCR conditions were as represented in Table 2.13.

Table 2. 13 Shows COBRA assay conditions.

Amplification	5' Primer (Forward)	3' Primer (Reverse)	PCR conditions
222bp of HERV-K LTR	COBRARHA	COBRARHB	96°C 30 sec (92°C 15 sec,50 °C 30 sec,68°C 30 sec) x34 cycle, 68°C 10 min,15 °C forever

2.11.3 Clean-up of PCR products

After running 3 µl of PCR products on an agarose gel to confirm the size and specificity of the product, the Cycle-Pure kit from Omega was used to clean-up the rest of PCR product, according to the manufacturer´s instructions. Purified PCR products from converted DNA was ready for direct sequence and / or cloning.

2.12 Genome –wide Amplification of Proviral sequence (GAPS) assay

The GAPS method was used as previously described in Macfarlane & Badge, (2015). This method was used to investigate insertions of HERV-K proviruses. The GAPS assay is designed to isolate only HERV proviral loci, and not amplify solo LTR loci. In this project, the HERV-K 3′LTR variant (3′ GAPS) was used exclusively. This assay included steps as follows:

2.12.1 DNA Digestion

For the construction of duplicate 3′ HERV LTR GAPS libraries, genomic DNA of melanoma cells lines (A375P, A375M, SKMel5 and SKMel28) and prostate (LNCAP), breast (T47D) and colon (SW480) cancer cell lines, were compared with the malignant pluripotent embryonal carcinoma NTera2D1cell line as a positive control were used. 600ng of genomic DNA was digested to completion in 1x NEB3 buffer and 20U *Ase* I enzyme, in a total volume of 30 µl.

The reaction was incubated overnight in a static water bath at 37^{o}C. Two controls were included: one without enzyme (replaced with an equal volume of 50% glycerol) and one without DNA (replaced with an equal volume of water). After complete digestion, the restriction enzyme was inactivated by heating to 65°C for 20 min.

2.12.2 Ligation

20µl of 20µM linker RBMSL2 and 20µl of 20µM RBD3 were mixed in equal volumes and heated for 10 min at 65°C, and then allowed to cool to room temperature over 30 min, to form the annealed linker. Each ligation was set up with 100 ng of the digested DNA and 2.7 µl of the 20µM annealed linker. Ligation was carried out overnight in a static water bath at 14 °C to 16°C in a final volume of 20 µl, which comprised digested DNA, annealed linker, 1x ligase buffer (NEB), 4 U of T4 DNA ligase enzyme and water to a volume of 20 µl. Ten out of 13 reactions included the restricted DNA. While three reactions were control reactions, including one reaction without ligase, one reaction without linker and one without DNA. After ligation the reactions were heated at 65°C for 20 min to inactivate the ligase enzyme.

2.12.3 Purification

Excess linkers and reaction component were removed by passing the ligation through a Microcon YM100 Centrifugal Filter Device, according to the protocol of the manufacturer. The retained ligated DNA was eluted in 20 µl TMT. The purified ligated DNA was aliquoted and stored at -20°C.

2.12.4 Suppression PCR (First PCR)

The reaction was set up in 20µl final volume, with 5µl of the ligated genomic DNA, 11.1x PCR buffer, 0.02 units / µl Taq DNA polymerase and primers (0.5 µM). CMK ENV (1.25µM) which was used for specifying the 3' end of the HERV LTR and the linker specific primer was RBX4. This PCR experiment also contained a “hood negative control” where water instead of DNA was added in the PCR hood and a “bench negative control” where water instead of DNA was added on the bench and the PCR conditions are shown in Table 2.14.

Table 2. 14 Suppression PCR assay conditions.

Amplification	5' Primer (Forward)	3' Primer (Reverse)	PCR conditions
HERV-K 3′LTR	CMKENV	RBX4	96°C 1 min (96°C 30 sec, 60 °C 2 min, (34 cycle), 59°C 10 min, 15 °C forever

2.12.5 Dilution step

Sufficient Single Molecule Dilution Diluent (SMDD= 5mM Tris HCl pH7.5, 5ng/ul sonicated *E.coli* genomic DNA) was prepared for all primary PCRs and controls (including secondary bench negatives). 98ul SMDD was aliquoted in thin wall PCR tubes in PCR clean racks. The secondary PCR mix was prepared and aliquoted at the same time. 2ul of primary PCR products were added into each 98ul of SMDD.

2.12.6 Nested PCR (Secondary PCR)

Secondary PCR mix was prepared for all reactions. The standard secondary PCR was 10 µl l in final volume, 9 µl mix + 1ul diluted (2+98ul) primary PCR. The master mix included, 11.1 x PCR buffer, 0.4 units Taq DNA polymerase and primers (1.25 µM). CMK3LTR was used to specify the 3' end of the HERV LTR and the linker was amplified with the primer RBY4. This PCR experiment also contained a “hood negative control” where water instead of DNA was added in the PCR hood and a “bench negative control” where water instead of DNA was added on the bench. The standard secondary cycle was the same as the primary PCR.

After that, 5 µl of secondary PCR products were run on 2% agarose gels at 120 V until all running samples were at the end of gel tray. Southern blotting was carried out as explained above. The remaining 5 µl was used for Next Generation Sequencing library construction.

2.12.7 Size selection of secondary PCR products

To size fractionate DNA fragments prior to use in the preparation of NGS libraries, samples were loaded on a 3% agarose gel and run at 120 V until all DNA bands were separated clearly. The gel was then transferred onto a Dark Reader visible light transilluminator and was cut to divide each sample into two different size ranges: more than 400bp (>400), and less than 400bp (<400). Each size of fragment of gel was transferred individually to a clean 1.5 ml DNA Lo-Bind Eppendorf tube. Then the DNA was extracted from the gel using the Zymoclean™ Gel DNA Recovery kit according to the manufacturer's instructions, as above. Eluted DNA concentration was measured by using a Nanodrop UV spectrophotometer method and diluted to get a DNA mass of 55ng / µl

2.12.8 Next Generation Sequence (NGS) library construction

The method of NGS library construction has four consecutive steps, which include fragmentation, ligation to barcoded adaptors, size selection and quantification of the library using the high sensitive DNA kit on an Agilent® 2100 Bioanalyzer.

2.12.9 Fragmentation of more than 400bp fragments (>400bp)

The Ion Xpress™ Plus Fragment Library kit protocol was used for the large size fraction (>400 bp) libraries. Briefly, the reaction was carried out using half of the volume of each reagent to reach a final volume of 20 µL. 55 ng of DNA was mixed with 2.5 µl of Ion Shear™ Plus 10X reaction buffer and nuclease-free water to bring the volume to 20 µl in a clean 1.5 ml DNA Lo-Bind Eppendorf tube. Then, 5 µl of Ion Shear™ enzyme mix II was added to the mixture for a total volume of 25 µl.

The reaction was mixed thoroughly about 8-10 times by pipetting, without generating any air bubbles. The tube was then incubated at 37°C for 8 min in a water bath to obtain a 350-450 base-read library. After that, 5 µl of Ion Shear™ stop buffer was added to the mixture, which was vortexed vigorously for at least 5 seconds.

2.12.10 Purification of fragmented DNA with Agencourt® AMPure® XP reagent

The Agencourt® AMPure® XP reagent beads were brought to room temperature and vortexed for 30 sec thoroughly to disperse the beads before use. Thereafter, 1.8 x sample volume of AMPure® XP Reagent was added to each prepared library. The mixture was then pipetted up and down 5 times to thoroughly mix the bead suspension with the DNA. The mixture was incubated for 5 minutes at room temperature. The tubes were placed in a magnetic rack and incubated for 2 minutes or until the solution cleared. The supernatant was carefully removed and discarded without disturbing the pellet. 500 µl of freshly prepared 70% ethanol was added and the tube was moved round while still on the magnetic rack to wash the beads, allow incubating for 2 min or until the solution clears. The supernatant was removed and discarded without disturbing the pellet. Those two steps were repeated for a second wash. It was ensured that all ethanol droplets were removed from the tubes. The tubes were kept in the magnetic rack, and the beads allowed to air dry at room temperature for 5 min. The tubes were rotated in the magnetic rack to assess the dryness of the bead pellet. To maximise the recovery of the DNA from the beads, it was important not to over-dry the pellet.

The tubes were removed from the magnet and 25 µl of pure water was added to the pellet to disperse the beads. The reactions were mixed by pipetting up and down 5 times and then the tubes were placed back on the magnet for at least 5 minutes.

The supernatant was removed and transferred to a clean 1.5 ml DNA Lo-Bind Eppendorf tube. Libraries were quantified using the high sensitivity DNA kit, according to the manufacturer´s instructions. Libraries were stored at −20°.

2.12.11 End repair assay for less than 400bp fragments (< 400bp)

The protocol was used for the fragmented (<400 bp) Libraries. Briefly, the reaction was carried out using half of the volume of each reagent. The 55 ng of DNA was mixed with 10 µl end repair 5x reaction buffer and 0.5 µl end repair enzyme was added to the mixture, with nuclease-free water to bring the volume to 50 µl in a clean 1.5 ml DNA Lo-Bind Eppendorf tube. Then, the tube was then incubated at 37°C for 20 min at room temperature. After that, Agencourt® AMPure® XP reagent beads were used to capture and purify the end repaired DNA, as explained above.

2.12.12 Ligate and nick-repair

The purified DNA fragments were then ligated to adaptors and nick-repaired. The reaction was carried out using half of the volume of each reagent by mixing 12.5 µl of the purified DNA with 5 µl of 10X ligase buffer, 1 µl of Ion P1 adaptor, 1 µl of Ion Xpress™ Barcode X, 1 µl dNTP mix, 24.5µl of nuclease-free water, 1 µl DNA ligase, and 4 µl nick repair polymerase in 0.2 mL PCR tubes.

The tubes were placed in a thermocycler at 25°C for 15 min, then at 72°C for 5 min and the samples were finally held at 4°C. Following the run on the thermocycler, the ligation reactions were purified with 1.2x volumes of Agencourt® AMPure® for libraries derived from amplicons greater than 400bp, while 1.5x for those derived from amplicons less than 400bp. The reactions were purified as explained above and their quality checked by use of the high sensitivity DNA kit on the Agilent® 2100 Bioanalyzer.

2.12.13 Size selection of DNA Library

Initially, the purified ligated libraries were separated by electrophoresis 2% agarose gel. The libraries derived from fragmented amplicons >400bp was cut between 200bp-500bp, while libraries derived from amplicons <400bp was cut between 100bp up to less than 500bp from gel, under the Dark Reader visible light transilluminator. Then the library DNA was extracted from the gel using the Zymoclean™ Gel DNA Recovery kit according to the manufacturer's instructions, as explained above.

The quantification of size selected library from each sample was performed using the high sensitivity DNA kit on an Agilent® 2100 Bioanalyzer. Three serial dilutions of each library were analysed by the bioanalyzer and the size distribution and molar concentration determined. Then the size selected library of each sample was pooled in equimolar amounts to make two pooled libraries, one from amplicons <400bp and another to from amplicons >400bp.

After pooling the libraries, the reaction tubes were handed onto to Rita Neumann (Department of Genetics and Genome Biology) for completion of the ion torrent sequencing protocol.

2.13 RNA Seq assay

This experiment aimed to discover differences in the transcriptome in different cell lines, growth conditions and phenotypes. In this experiment, RNA from SKMel28 and SKMel5 grown in 10% FBS media was compared to RNA from SKMel5 under 1% FBS culture, showing the suspension phenotype as well as SKMel28 which maintained the adherent phenotype in 1% FBS . In addition, RNA from normal human, melanocytes

were used as a control. This experiment was done with three biological repeats and two technical replicates per sample.

The lllumina TruSeq stranded total RNA sample preparation low sample (LS) protocol was used according the manufacturer's instructions. The workflow of the RNA seq experiment is illustrated in Figure 2.5.

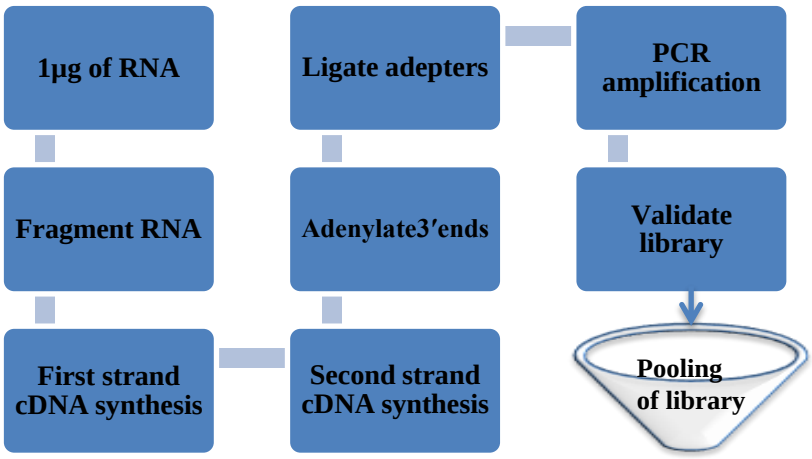


Figure2.53: Preparation of cDNA and library preparation for sequencing

The workflow shows process of RNA to synthesis cDNA. The quality of RNAs was checked using the Agilent bioanalyser 2100 and only samples with an RNA integrity value (RIN) of greater than 9 were used for this experiment.

2.13.1 Ribo-zero rRNA depletion and RNA fragmentation

1 µg of RNA was used, diluted in nuclease-free water to a final volume of 10 µl. Diluted RNA samples were transferred to a clean 96-well plate. Then, 5µl rRNA Binding Buffer was added to each well of the plate. 5µl of rRNA. Removal Mix reaction was added to each well and mixed gently by pipetting 6 times. The PCR plate was then incubated in a thermal cycler at 68°C for 5 min, after which the plate was incubated at room temperature for 1 minute.

35µl of rRNA removal bead solution was added to each well. Then, 20 µl was transferred from each well to a new plate. The PCR plate was incubated at room temperature for 1 min and then put onto a magnetic stand, at room temperature, for 1 minute. The supernatant was then transferred to a new 96-well PCR plate. The next step, clean-up of the RNA, was done according to the manufacturer´s instructions.

2.13.2 Synthesizing First Strand cDNA

1µl of SuperScript II enzyme mix was added to 9 µl of first strand synthesis buffer and mixed gently. 8 µl of this mixture was added to each well of the 96 well plate containing the rRNA depleted RNA was gently pipetted 6 times to mix thoroughly. The plate was then incubated in a thermal cycler at 100°C. The PCR program is 25°C for 10 min, 42°C for 15 min, 70°C for 15 min and then held at 4°C.

2.13.3 Synthesizing Second Strand cDNA

To synthesis second strand cDNA, the end repair control solution was diluted to 1/50 in resuspension buffer. 5µl of end repair control solution was added to each well. After thawing and centrifugation of the second strand synthesis master mix, 20 µl of second strand synthesis master mix was added to each well and gently pipetted 6 times to mix thoroughly. The plate was incubated in a thermal cycler at 16°C for 1 hour. Clean-up was repeated by The Agencourt® AMPure® XP reagent beads.

2.13.4 Adenylate 3'Ends

A-tailing control was diluted 1/100 in resuspension buffer. Then, 2.5 µl of the diluted A-tailing control was added to each well of the PCR plate. 12.5 µl of A-tailing mix was added to each well and pipetted 10 times to mix thoroughly. The plate was incubated and set up a thermal cycler at 100°C. The PCR program is 37°C for 30 min, 70°C for 5 min and held at 4°C.

2.13.5 Ligation of Adapters

The ligation control was diluted to 1/100 in resuspension buffer and 2.5µl of diluted ligation control was added to each well of the plate. Then, 2.5µl of ligation mix was added to each well. RNA adapter solution 2.5µl was then added and pipetted 10 times to mix thoroughly. The plate was incubated in a thermal cycler at 30°C for 10 min. After incubation, 5 µl of stop Ligation buffer was added to each well of the plate, to inactivate the ligase, and pipetted 10 times to mix thoroughly. Clean up was repeated again.

2.13.6 Enrichment of DNA Fragments

5 µl of PCR primer cocktail was mixed with 25 µl of PCR master mix and added to each well of the PCR plate. This mixture was gently pipetted 10 times to mix thoroughly. The PCR plate was then incubated in a thermal cycler at 98°C for 30 sec , then 15 cycles of: 98°C for 10 sec , 60°C for 30 sec ,72°C for 30 sec, 72°C for 5 min and held at 4°C. The clean-up was repeated.

2.13.7 Quality assessment of libraries

The quality assessment of the library was done by using the high sensitivity DNA kit on the Agilent 2100 Bioanalyser. The size of the library should be a band centred at approximately 260 bp.

2.13.8 Quantification of libraries

To check the library quality prior to use on the illumina sequencing platform, two methods were use. The first method involved qPCR using the KAPA library quant kit carried out according to the manufacturer's instructions. The second method, the Quanti fluor® dsDNA system from Promega was used as instructed by the manufacturer. After that, the libraries were pooled and sent to DeepSeq at the University Nottingham for sequencing on the illumina NextSeq instrument.

3 Quantification of HERV K expression in cancer cell lines

3.1 Introduction

HERV-K is the most recently integrated endogenous virus in the DNA of humans, and retains the ability to express viral proteins (Bhardwaj *et al.*, 2015a). Many reports have shown that HERV-K genes are significantly upregulated in melanoma, prostate, breast and colon cancers (Serafino *et al.*, 2009; Schmitt *et al.*, 2013; Lee *et al.*, 2014; Bhardwaj *et al.*, 2015; Goering *et al.*, 2015).

In a previous study by (Serafino *et al.*, 2009) it was shown that under stress conditions, such as serum deprivation (only 1% FBS in the culture medium), the melanoma cell line TVM-A12, which normally grows as adherent cells in monolayer, changes its phenotype. Stress induced modification of morphology involves a change to suspension growth, in clusters of surface detached cells. In this state these cells show a higher proliferation rate, and a significant increase in human endogenous retrovirus (HERV) gene expression in comparison to adherent TVM-A12. It was observed that HERV-K *env* gene expression, detected by real-time PCR, was highest in adherent TVM-A12 grown in 1% FBS compared to 10% FBS.

Following the Serafino *et al*, (2009), this study explored how a variety of other melanoma cell lines, as well as other cancer cell lines, responded to serum starvation with respect to HERV-K gene expression. Our project started by selecting four melanoma cells (A375P, A375M, SKMel5 and SKMel28) and three other cancer cell lines (LNCAP (Prostate), T47D (Breast) and SW480 (Colon)).

In this chapter, all cancer cells were analysed in two conditions: 10% FBS and 1% FBS. The expression of HERV-K *pol* mRNA was investigated in the melanoma cell lines and other cancer cells, with the aim of corroborating previously published findings and establishing whether this was a feature of all melanoma cell lines, or just the TVM-12

cell line. It expected to identify an increase HERV-K *pol* gene expression in 1% FBS culture in melanoma cell lines. Additionally, we determined HERV-K *pol* gene expression in normal primary melanocytes as a control for melanoma HERV-K expression.

3.2 Results

3.2.1 Profile of cancer cells

Cell line authentication and cross-contamination testing are essential steps to confirm the reliability of experiments using cell lines. There are different sources of cell line contamination, such as mixing of cell lines, mislabelling, particularly if these cells have the same morphology in culture (Reid *et al.*, 2004). The seven cancer cell lines selected for analysis (A375P, A375M, SKMel5, SKMel28, LNCAP, SW480 and T47D) were sent to for DNA identification by Genetica DNA Laboratories (Burlington NC).

Cell line authentication was done by genetic profiling using polymorphic short tandem repeat (STR) loci. These characteristic repeats can be used for the detection and identification of contaminating human cells, and to establish the provenance of cell cultures (Reid *et al.*, 2004).

The DNA profile of all the cancer cell lines used were identical to the American Type Culture Collection (ATCC) data, with one exception in A375P (Table 3.1). Interestingly there were differences between the A375P and A375M cell lines.

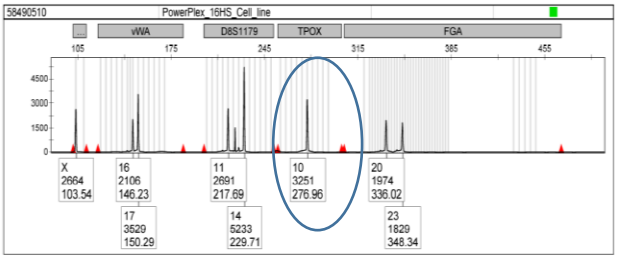
These are clonal derivatives of the same original melanoma, with A375P is a poorly metastatic melanoma cell line, as judged by lung metastasis formation in nude mice (Xu *et al.*, 2008). A375M is a clonal derivative of A375P, derived from a nude mouse lung metastasis by Xu *et al*. (2008), which is metastatic in these assays. Interestingly, the 8

repeat allele at TPOX was missing from A375P when compared with A375M, as shown in Figure 3.1. These data suggest that the culture of A375P used in this thesis has undergone loss of heterozygosity at TPOX, while A375M retains the expected DNA fingerprint.

Table 3. 1 DNA profile of cancer cells by STR. The loss of an allele at TPOX in A375P is highlighted in blue.

Name of cells / Name of markers	A375P	A375M	SKMel5	SKMel28	LNCAP	SW480	T47D
D3S1358	15,17	15,17	16,17	16,18	16,17	15	15,17
D7S820	9	9	9,12	9.3,10	9,9.1,10.3	8	11
vWA	16,17	16,17	14,18	16,19	16,18	16	14
FGA	20,23	20,23	20.2,22	19	19,20	24	18,23
D8S1179	11,14	11,14	12,15	13	12,14	13	13
D21S11	29,30	29,30	29	28,29	29,32,2	30,30.2	28,31
D18S51	12,17	12,17	15,16	12,16	11,12	13	17
D5S818	12	12	11,13	11,13	11,12	13	12
D13S317	11,14	11,14	10,12	11,12	10,12	12	12
D16S539	9	9	10,12	9,12	11	13	10
TH01	8	8	6,9	7	9	8	6
TPOX	10	8,10	11	8,12	8,9	11	11
CSF1PO	11,12	11 ,12	10,13	10,12	10,11	13,14	11,13
AMEL	X	X	X	X	X	X	X
Penta D	9,15	9,15	9.11	9,10	12,12.4	9,15	10,12
Penta E	10,12	10,12	5,12	8,12	12,16	10	7,14

A375P



A375M

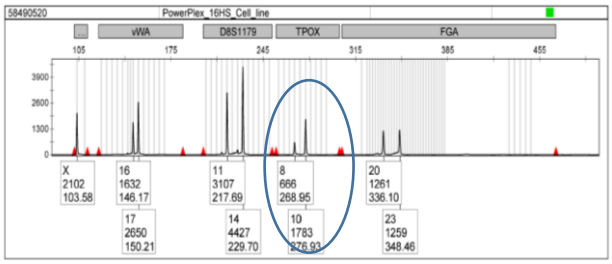


Figure 3.1: Electropherogram of A375P and A375M cell lines.

A375P at TPOX locus has lost one allele, while A375M has two alleles (highlighted by the blue oval).

3.2.2 TERT gene promoter mutation in A375P and A375M cell lines

Generally, two mutations within the TERT promoter C228T and C280T occur in melanoma and other cancer cell lines (Huang *et al.*, 2013), leading to TERT reactivation. To further investigate the genotype differences between A375P and A375M analysis of the TERT promoter was set up. As A375P and A375M are clonal derivatives we expected them to carry the same TERT activating mutation. To detect whether this was the case, we sequenced the region of the TERT promoter containing the mutations, in A375P and A375M.

3.2.2.1 Processing of PCR and data analysis

PCR reactions were set up with TERTC228-2509 (forward) and TERTC228-250 (reverse) primers encompassing the mutated region. The pilot experiment used the DNA of the A375P melanoma cell line. The experiment included a negative control (where water replaced DNA), as illustrated in figure (3.2).

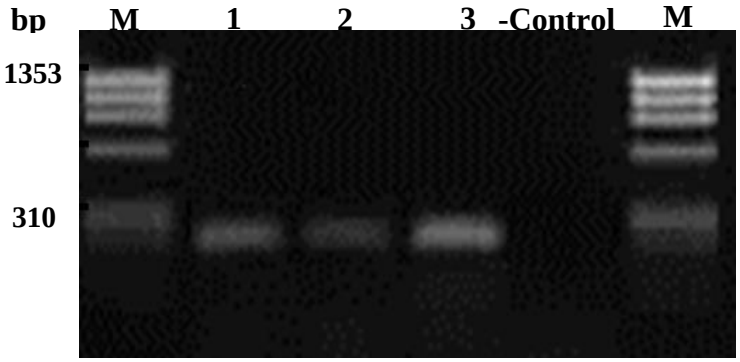


Figure 3.2: PCR with DNA from A375P cell line.

DNA from three different passages (1-3) of the A375P cell line were used with TERTC228-250(forward) and TERTC228-225(reverse) primers. The product was less than 300bp, as predicted. ϕX174 *Hae III* marker (M) was used. The negative control (-) did not show any amplification.

The PCR product was excised, frozen and thawed to provide an eluate for PCR re-amplification (20ul), for sequencing. The re-amplified PCR products were run out on a window gel (see Methods and Materials) and the bands pipetted up when they migrated into the window. The DNA was purified and the concentration determined by UV spectrophotometry.

Two sequencing reactions were set up, one with TERTC228-250 (forward) primer while the other used the TERTC228-250 (reverse) primer. The reactions gave good sequencing results with both primers, confirmed by using BLAT searches on the human reference genome at the UCSC Genome Browser website.

These steps were repeated with DNA from the A375M cell line. Then the sequence data from the A375P and A375M cell lines was aligned to the human genome reference sequence using the Clustal Omega program at the EBI website www.ebi.ac.uk/Tools/msa/clustalo/, to locate the mutation in the cell lines. The mutation in C228T position was observed in both forward and reverse sequencing reaction, in both cell lines, while the mutation at C250T mutation was not observed, as illustrated figure (3.3).

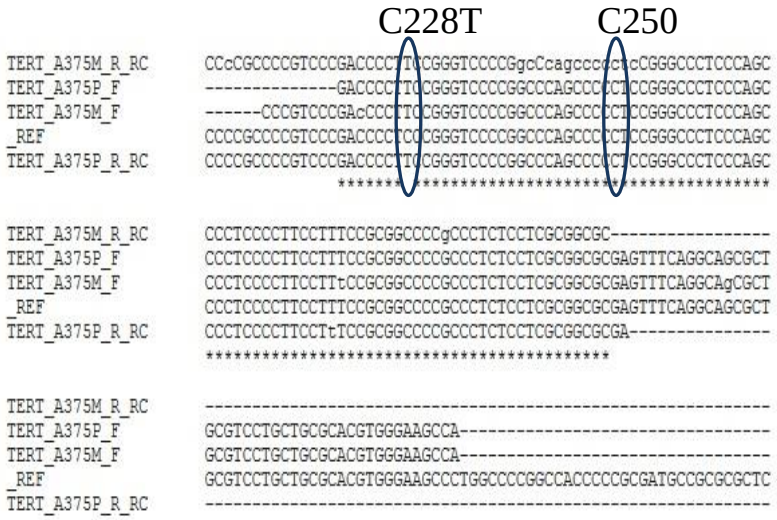


Figure 3.3: Analysis of sequencing of A375 cell line.

Four strands of A375P and A375M sequence data aligned with the human genome reference sequence. The mutation is highlighted in blue oval at the top of the sequence C228Tposition. However, the mutation at the C250T position was not observed.

In conclusion, a number of differences in the genotype of the A375P andA375M cell lines were observed. This clearly pointed out these cell lines were not genetically identical and may have different patterns of gene expression and epigenetic state.

3.2.3 Phenotypes of cancer cell lines under two culture conditions

3.2.3.1 Phenotypes of cancer cell lines culture under serum starvation

To determine if melanoma cells (A375P, A375M, SKMel5 and SKMel28) and other cancer cell lines (LNCAP, SW480 and T47D) would change their phenotype and HERV-K expression under stress, cells were grown in two conditions 10% FBS and 1% FBS for 4 days. All 7 cancer cell lines were cultured in duplicate flasks: the first flask included 10% FBS, normal conditions for growing human cells, while the other contained media with only 1% FBS. Over two weeks, the cultures were observed every day to detect any changes in phenotype. Cellular morphology was recorded by microphotography.

SKMel5 was observed to adopt a suspension phenotype (SUS) after 48 hours and A375M slightly changed its phenotype. However, the change was not observed in cultures of SKMel28, T47D, LNCAP, SW480 and A375P. At the same time, all cancer cells lines were adherent (AD) in 10% FBS. Therefore, our data showed that not all melanoma cell lines change their phenotype under stress conditions to become suspension cells, as observed by Serafino *et al.*, for the melanoma cell line TVM-A12 (Figure3.4).

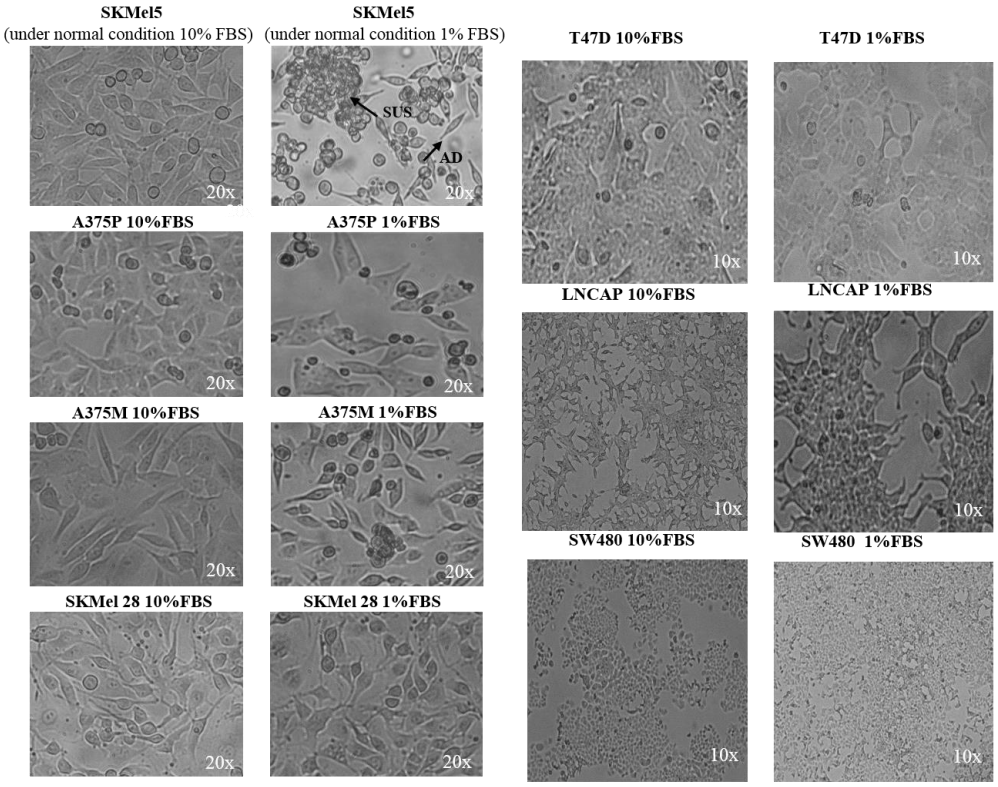


Figure 3.4: Cancer cell lines under two culture conditions.

Phase contrast images were taken with 20x and 10x objectives after 4 days of culture. Some lines especially SKMel5 showed clusters of cell and rounded shape the black arrows indicate suspension (SUS) and adherent (AD) phenotype. In contrast, when SKMel5 was grown in normal conditions (10% FBS) it form adherent layer. Similarly, A375M showed subtle changes in phenotype, with a few clusters of suspended cells. However, the modification of the adherent phenotype was not shown in cultures of A375P, SKMel28, SW48, LNCP and T47D.

3.2.3.2 Validation of growing cancer cells in stress condition 1%FBS

3.2.3.2.1 Measuring proliferation of adherent cancer cells under 1% FBS by colony formation

We further determined how the stress condition of culture in 1% FBS containing media impacts on the growth of cancer cells by using colony formation assays. This assay allows us to confirm these cells were growing normally in this condition, or not.

Melanoma and other cancer cells were cultured in 10 % FBS and 1% FBS containing media. 1000 cells / well and 2000 cells / well were plated, and incubated for 7 days.

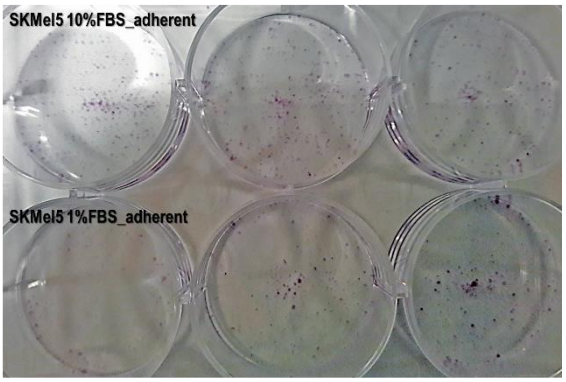
The number of colonies of cells were counted by ImageJ and manually to know exactly the numbers of colonies cells growing in the two conditions, as illustrated in Figure 3.5.

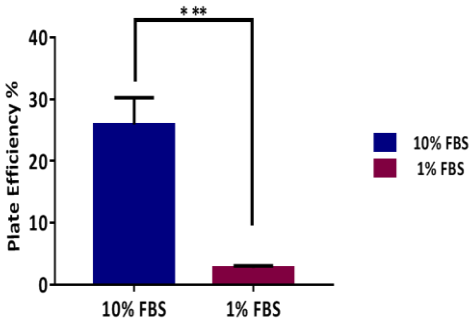
It is clear that the numbers of colonies of SKMel5 and SKMel28, and other cancer cells LNCAP, SW480 and T47D were drastically reduced in1% FBS compared to the control (10% FBS), but that the cells were still viable and growing (The data in Appendix 2).

By contrast, A375P and A375M colonies were completely removed after the fixation and staining steps (Figure 3.6). This might suggest that these cells were not able to form colonies under 1% FBS, despite the robust growth observed previously (Figure 3.4). However microscopy of the colony formation assay plates showed colonies on the plate surface that were lost on fixation (Figure 3.6 insert). This strongly suggests that although theA375 lines do not show complete reversion to suspension growth, their ability to adhere to culture dish surfaces is impaired by serum starvation.

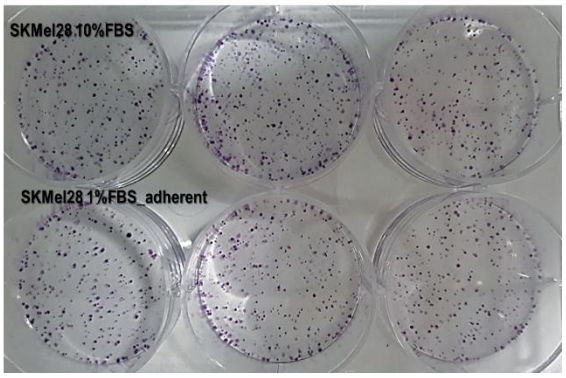
These results raised the possibility that the apparent suspension phenotype, seen in the SKMEL5 cell line, was a result of the loss of cell surface adhesion molecules, leading to detachment and cell death. As a result, we needed to investigate whether the observed suspension cells, especially SKMel5 (as mentioned in Figure 3.4) were truly viable, or simply dead and dying cells that had detached. To test this hypothesis, a cell viability assay was setup because the colony formation assay is not suitable for suspension cells.

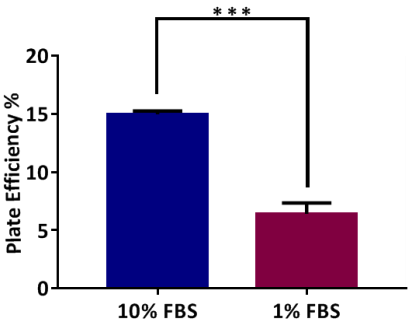
A


SKMel5 10%FBS_adherent
SKMel5 1%FBS_adherent

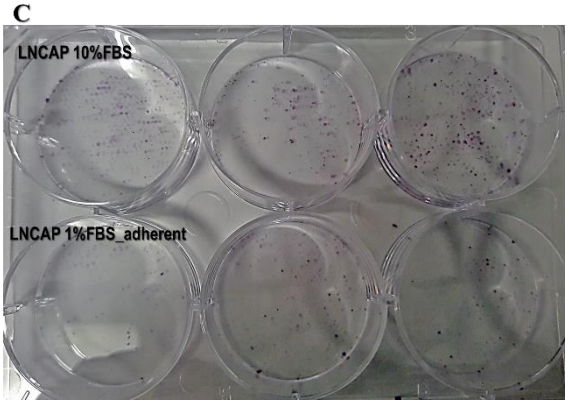

***
40
30
20
10
0
Plate Efficiency %
10% FBS
1% FBS
10% FBS
1% FBS

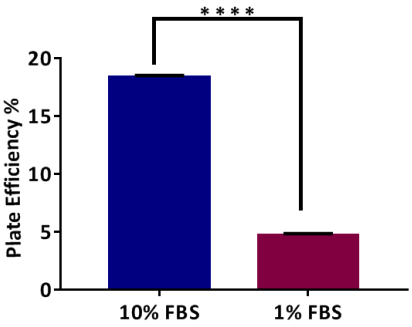
B


SKMel28 10%FBS
SKMel28 1%FBS_adherent


***
20
15
10
5
0
Plate Efficiency %
10% FBS
1% FBS

C


LNCAP 10%FBS
LNCAP 1%FBS_adherent


****
20
15
10
5
0
Plate Efficiency %
10% FBS
1% FBS

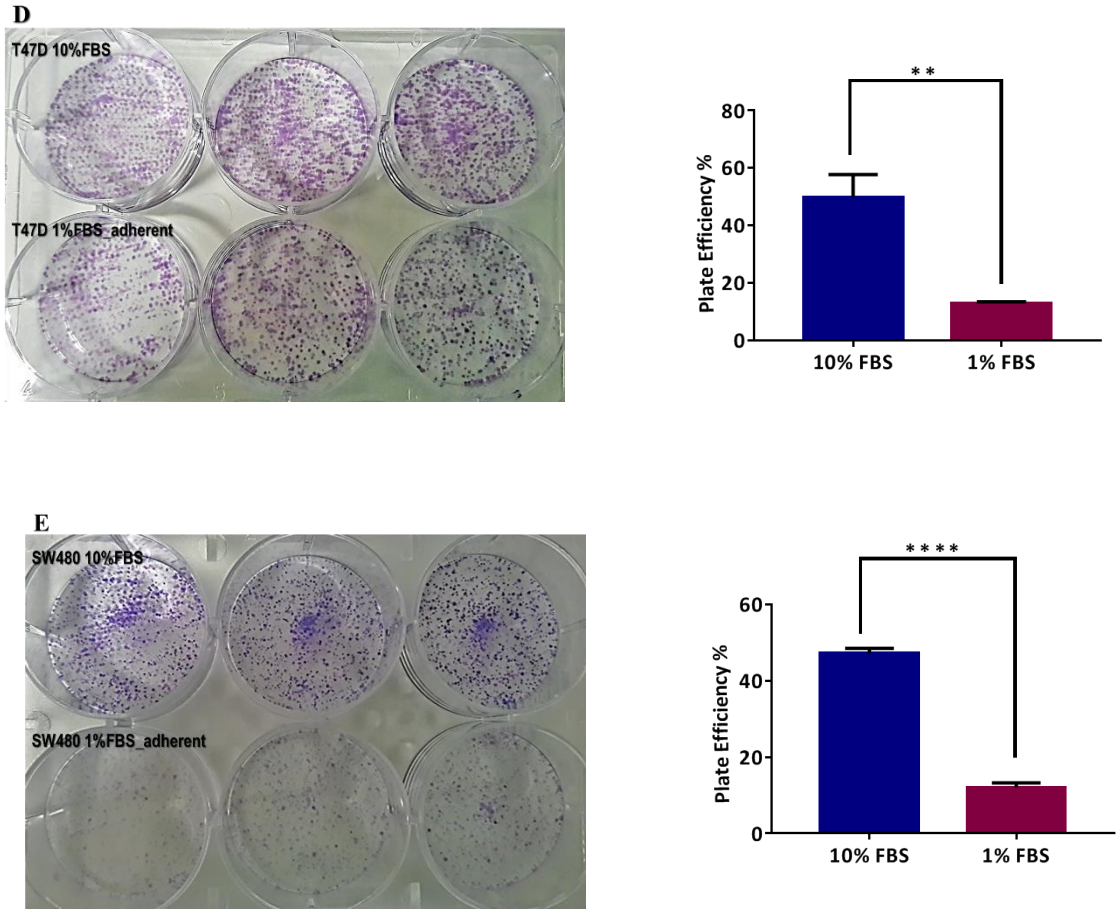


Figure 3.5: Determination of plating efficiency of melanoma cells and other cancer cells under two conditions.

The cells in 10% FBS were seeded at 1000 cells / well and the cells in 1% FBS were seeded at 2000 cells / well in 6 well plates. After 7 days, the plates were stained for the formation of cell colonies with crystal violet dye. The picture of the colonies was then taken using a digital camera, and the colony numbers counted. The cells under 10% FBS showed significantly higher plating efficiency (colony formation rate) compared to those in 1%FBS. The values plotted are means ± SEMs from three biological repeats and three repeats (Student's unpaired t test), * p< 0.05, ** p< 0.01, *** p< 0.001, p< 0.0001 vs Control. A: SKMel5, B: SKMeL28, C: LNCAP, D: T47D, E: SW480

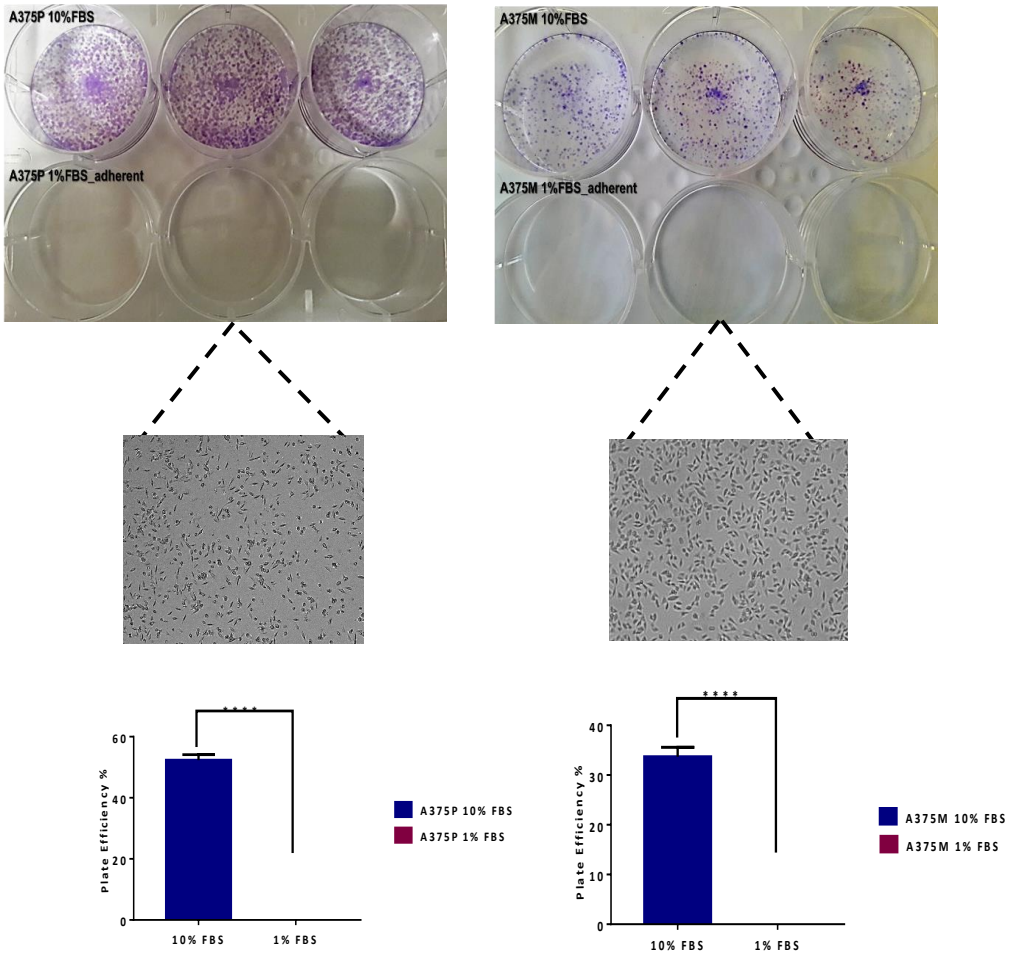


Figure 3.6: Determination of growth of A375P and A375M cells under two conditions.

A375P andA375M were seeded in 10%FBS and 1% FBS media. After 7 days, the plates were stained for the formation of cell colonies with crystal violet dye. However, both strains of cells were removed from 6 well plates under 1% FBS after the process of this assay when compared with 10% FBS. Pictures of the colonies were then taken using a digital camera before starting the colony formation assay, and the colony numbers in 10% FBS were counted. The cancer cells under 10% FBS were recorded high significantly compared to 1% FBS. The values plotted are means ± SEMs from three biological replicates and three repeats (Student's unpaired t test), * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ vs Control.

3.2.3.3 Measuring cell proliferation under 1% FBS by viability assay.

To determine the metabolic activity of the suspension phenotype cells, in 1% FBS culture conditions, the Promega RealTime-Glo™ MT Cell Viability. The viability assay was set up with 4 end points hours (h) (1h, 24h, 48h and 72h). We analysed SKMel5 (mostly suspension cells), A375P (adherent) and A375M (small clusters of suspension cells). As seen in Figure 3.7. These cells grew in 1% FBS, showing adherent and suspension phenotype, and in 10% FBS as a control. In addition, media alone was used as a negative control in this experiment.

At the start of the assay (0 h), SKMel5, A375P and A375M melanoma cell lines showed the same level of growth under 10% FBS, but A375P and A375M showed much reduced metabolic activity in 1% FBS (Figure 3.7 B and C, far left). However, after 24h all cells showed significantly reduced metabolic activity under 1% FBS (Figure 3.7 A / B / C). Both A375P and A375M showed significantly higher metabolic activity in the remaining adherent cells, suggesting that the cells suspended in the media, were dying, or at least no metabolically active. In sharp contrast, at the start of the experiment adherent and suspension cells of the SKMel5 cell line showed no significant difference in metabolic activity.

After 24 hours although overall cell activity was reduced, but the SKMel5 suspension cells were significantly more active that adherent cells (Figure 3.7 A) this trend continued at the 48 and 72 hour time points. These results show that the SKMel5 cell line is robust to serum starvation and the high proportion of suspension cells observed results from actively growing cells that have detached, rather than a high rate of cell death. The data also suggest that the slight increase in suspension cells observed in A375M cultures is not due to viable cells detaching, but dead and dying cells in the media.

Taken together with the colony formation assay data these results indicate that loss of cell surface adhesion is seen in some melanoma cell lines (SKMel5, A375P and A375M), but not (SKMel28), and that in a minority of cases (SKMel5) some melanoma cell lines were able to transition to active growth in suspension.

Serafino *et al*., (2009) observed that changes of cell phenotype induced by serum starvation led to higher HERV-K expression, and that inhibition of HERV expression could revert the cellular phenotype, suggesting a direct link between cell phenotype and HERV-K expression. In the case of TVM-A12 suspension clones it was suggested that these changes were consistent with greater metastatic potential. To investigate this possibility in the SKMel5 line we next examined HERV-K gene expression by RT-PCR.

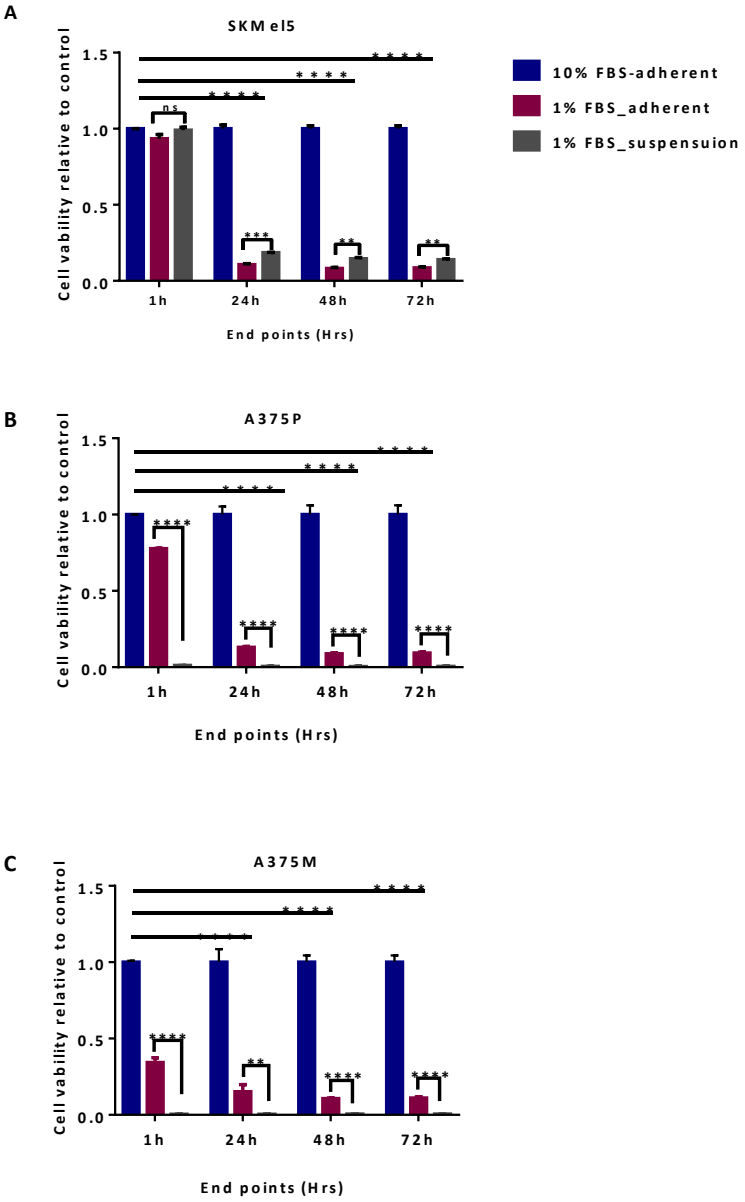


Figure 3.7: Measuring metabolic activity of melanoma cells under 1% FBS as suspension by viability assay.

Viability assay was done on three melanoma cell lines. 1500 cells were seeded in 6-well plates in 10% FBS and 1% FBS containing media. After 24 hours seeding, the viability reagents were added to cells. The viability cells was measured at four time points, 1% FBS as stress condition impacted on the growth of the cells. **(A)** SKMel5, **(B)** A375P and **(C)** A375M. Our data showed the SKMel5 suspension cells maintained viability under 1%FBS. The values plotted are means ± SEMs from three biological repeats and four repeats by two-way ANOVA, followed by Student's paired t test between two groups. Non-significant (ns),* p <0.05, ** p <0.01, *** p <0.001, **** p <0.0001 vs Control.

3.2.4 Detection of HERV- K *pol* gene expression in cancer cell line

3.2.4.1 Analysis of HERV-K *pol* gene expression in melanoma cells and other cancer cell lines by RT-PCR and qPCR

Gene expression from HERV-K HML2 proviruses is dominated by two alternatively spliced forms of the genomic transcript that share a 5'UTR and are initiated from the promoter in the 5'LTR. The *pol* protein is cleaved from a polyprotein comprised of the *gag-pro-pol* encoded peptides. The second spliced transcript encodes the *env* protein. As a result, the expression level of the gag and pol genes are expected to correlate closely, but env could be expressed at higher (or lower levels) depending the bias of splicing in favour of *gag-pro-pol* or *env* encoding transcripts. We extracted RNA from melanoma and other cancer cell lines under the two culture conditions (10% and 1% FBS) and checked the integrity and concentration of the RNA fragments using the Agilent RNA 6000 Nano Kit. After cDNA synthesis, specific primers were used to amplify the HERV-K *pol* gene. To check the cDNA was of high quality and the Reverse Transcriptase negative control (RT-) was clear of contamination, the RT-PCR reactions were fractionated by agarose gel electrophoresis (Figure 3.8). The experiment included two negative controls; the RT- control, where the RT enzyme was inactivated prior to cDNA synthesis and a PCR negative where the cDNA was replaced by water (–PCR). The PCR positive control was genomic DNA from A375M, which showed an identical size amplicon as from cDNA products, as HERV-K genes lack introns. As shown in Figure 3.8, different levels of expression of the *pol* gene was observed in melanoma and other cancer cell lines, but all showed some *pol* expression.

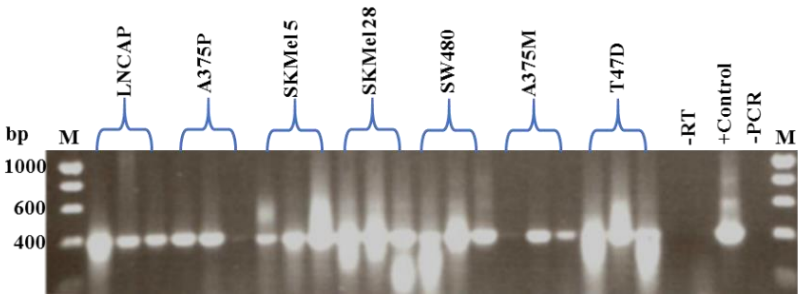


Figure 3.8: Expression of the *pol* gene in cancer cell lines.

The RT- PCR assay was performed to detect viral *pol* gene expression in normal culture conditions (10% FBS), in seven cancer cells (A375M, A375P, SKMel5, SKMel28, T47D, LNCaP, and SW480). 35 cycles amplification was used with the *pol* primers (IASPOL F and IASPOLR). The RT- PCR product produced was expected to be 400bp in length. The experiment included two negative controls (-RT,-PCR) to exclude DNA contamination and positive control (+) gDNA. Hyperladder 1 (H1) marker was used. This experiment was done with three biological replicates per cell line.

Next, we investigated expression of the *pol* gene in cancer cell lines by Quantitative Real time PCR (qPCR) reactions. The housekeeping genes Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), phosphoglycerate kinase 1 (PGK1) and succinate dehydrogenase complex flavoprotein subunit A (SDHA) were used as reference genes to normalise pol expression between culture conditions. cDNAs derived from 10% and 1% FBS (adherent cells) were used to measure *pol* expression in both conditions. Melanoma cell lines and other cancer cell lines were grown under stress conditions with no change in phenotype, with the exception of SKMel5, which changed from adherent to suspension phenotype. This experiment provided a means of quantitatively comparing *pol* expression between 10% and 1% FBS culture conditions. The results of this experiment are shown in Figure 3.9. A375P showed significantly increased expression of *pol* in 1% FBS compared to 10% FBS. However, A375M and SKMel5 showed non-significant differences in expression between the two conditions. By contrast, SKMel28 showed significantly reduced *pol* expression under 1% FBS, which was not expected.

Interestingly this experiment showed that changes in the expression of *pol* between A375P and A375M when starved of serum were not identical, indicating a big difference between these two almost genetically identical cell lines. Intriguingly A375P was the only line to show a significant increase in HERV-K pol expression under stress conditions, as was observed by Serafino *et al.*, (2009) for their melanoma cell line isolate.

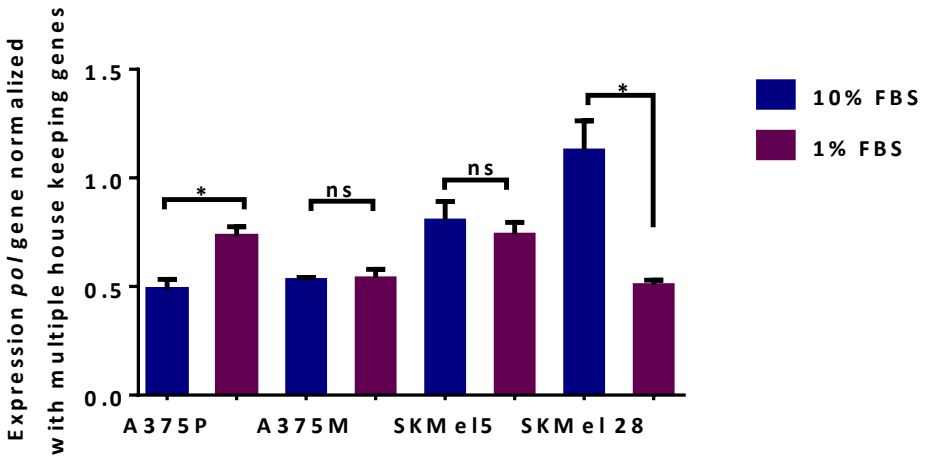


Figure 3.9: HERV-K *pol* gene expression in Melanoma cells.

The Applied Biosystems® Step OnePlus™ Real-Time PCR Systems with Fast SYBR® Green Master Mix was used to detect HERV K *pol* expression and to compare between adherent cells in 10% and 1%FBS conditions. The data showed that the melanoma cell lines had different levels of expression, especially A375P. The bars represent relative expression of the target gene (*pol*) normalized with multiple housekeeping genes, by following the Pfaffl method (Pfaffl, 2001). The values plotted are means ± SEMs from three biological repeats. Non-significant (ns),* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ vs Control. (Student's paired t test)

As it appeared that the upregulation of HERV-K *pol* expression on serum starvation was variable between melanoma cells lines, and not associated with change to a suspension phenotype, we next considered the effect of this treatment on other cancer cell lines. We measured changes in *pol* expression in LNCAP, T47D and SW480 between the two conditions by qPCR, normalizing with multiple housekeeping genes (Figure 3.10). It was observed that the SW480 colon cancer cell line had significantly higher HERV-K *pol* expression under 1% FBS, compared to *pol* expression in 10% FBS. While both LNCAP and T47D showed slightly increased expression under 1% FBS, neither effect was statistically significant (Figure 3.10).

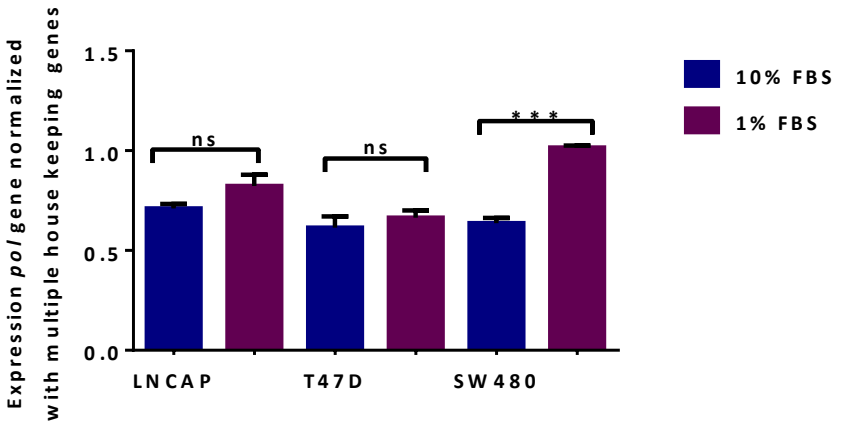


Figure 3.10: Quantitative HERVK *pol* gene analysis of other cancer cell lines.

Pol gene expressed in SW480 was greater than LNCAP and T47D under stress condition. Generally, the data showed that cancer cell lines analysed were increased *pol* expression under 1% FBS. The bars represent relative expression of the target gene (*pol*) normalized with multiple housing genes. The values plotted are means ± SEMs from three biological repeats. Non-significant (ns),* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ vs Control. (Student's unpaired T-test).

In addition, our results revealed the fold change was 5.6 in A375P more than in 10% FBS, while A375M did not change significantly in *pol* expression in either condition. In SKMel5 and SKMel28 pol expression was reduced -1.7 and -5.1 fold respectively.

In SW480, we found that level of *pol* was 3 fold higher significantly than in LNCAP and T47D (Figure3.11).

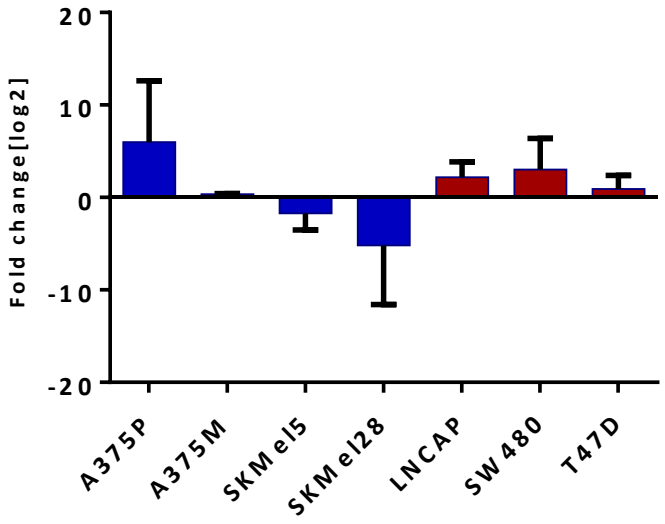


Figure 3.11: Quantitative HERVK *pol* gene expression, displayed as fold change.

Pol gene was expressed as fold change in melanoma cells and other cancer cells under 10% FBS, compared with 1%FBS. Bars (means ± SE) extending above the baseline show increased expression in A375P, LNCAP, SW480 and T47D, under 1% serum conditions and bars extending below the baseline show lower expression in SKMel5 and SKMel28 cells compared with 10%FBS. A375M did not show significantly changed expression. Results showing relative expressions of mRNA of the target gene (*pol*) normalized with multiple housekeeping genes and calculated as Pfaffl method by The REST384 software. The values plotted are means ± SEMs from three biological repeats. Blue colour is melanoma cells and red colour is other cancer cell lines (prostate, colon and breast) respectively.

3.2.4.2 Quantification of HERV-K *pol* gene expression in SKMel5 adherent and non-adherent (suspension) cell line under 1% FBS stress.

To investigate how the stress condition changed *pol* gene expression, we performed qPCR on cDNA from of SKMel5 cultures grown in in various conditions.

As this was the only cell line that showed a change from adherent to suspension phenotype it was possible to separate adherent and suspension cells, rather than analyse them together. As illustrated in Figure 3.12. SKMel5 showed different levels of *pol*

expression in 10 % FBS (adherent), 1% FBS (adherent) and 1% FBS (suspension) cells, although these differences between 10% FBS and 1% FBS adherent were not significant by Students T-test. However, 1%FBS SKMel5 suspension was significantly reduced expression compared to 1% FBS adherent culture. Nevertheless, SKMel5 cell grown in 10%FBS recorded higher expression compared with 1% FBS (adherent) and 1% FBS (suspension). This reduction in *pol* gene expression, on serum starvation and change from adherent phenotype, was only seen in the SKMel5 cell line, and disagrees with the result of Serafino *et al.*, 2009.

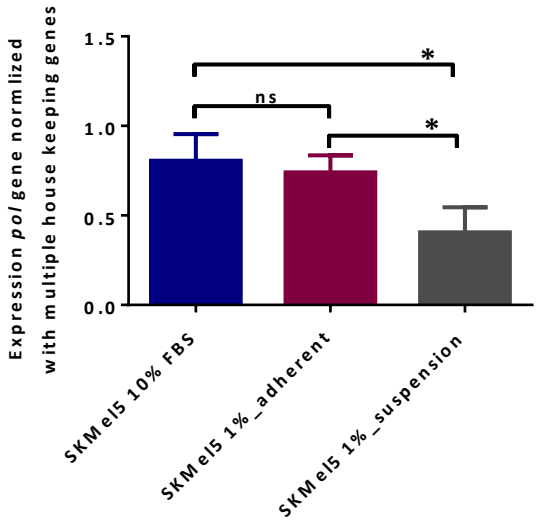


Figure 3.12: Quantitative HERV-K *pol* gene in SKMel5 adherent and non -adherent (suspension) cell line under 1% FBS.

The *pol* gene is expressed more in SKMel5 under 10% FBS than under 1% FBS. At the same time SKMel5 adherent cells expressed the *pol* gene more than cell with a suspension phenotype. Generally, the data showed that the expression of *pol* gene was gradually decreasing in serum starvation. The bars represent relative expression of the target gene (*pol*) normalized with multiple housekeeping genes. The values plotted are means ± SEMs from three biological repeats. Non-significant (ns),* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ vs Control. (Student's unpaired T-test).

3.2.4.3 Quantification of HERV-K *pol* expression in melanocytes

We were next interested in comparing the mRNA expression of HERV-K *pol* between our melanoma cell lines and primary melanocytes. We utilised the melanoma cell lines A375P, A375M, SKMel5 and SKMel28 as an *in vitro* model of melanoma. The melanocytes used were Primary human epidermal melanocytes (HEMaLP) isolated from adult skin. These cells are isolated from an individual with lightly pigmented tissue. All cell lines were grown in normal culture conditions (10% FBS). Then RNA was extracted from each of the cell lines and treated with DNAase to prevent genomic DNA contamination. This treated RNA was then converted to cDNA. cDNA was subsequently used in a qPCR reaction in order to relatively quantify the levels of each *pol* gene. This data was normalized relative to multiple housekeeping genes. Our results revealed that HERV-K *pol* gene was expressed in all melanoma cell lines and melanocytes (Figure 3.13).

It is clear however that the level of *pol* expression in the A375P and A375M cell lines was significantly lower than in melanocytes. By contrast the other two melanoma cell lines, SKMel5 and SKMel28 showed levels of *pol* expression that were not significantly different from that seen in melanocytes. According to these data, we can infer that the expression of HERV-K *pol* in melanoma and melanocytes is very similar, and it is difficult to see how HERV-K *pol* expression could be a reliable biomarker for melanoma.

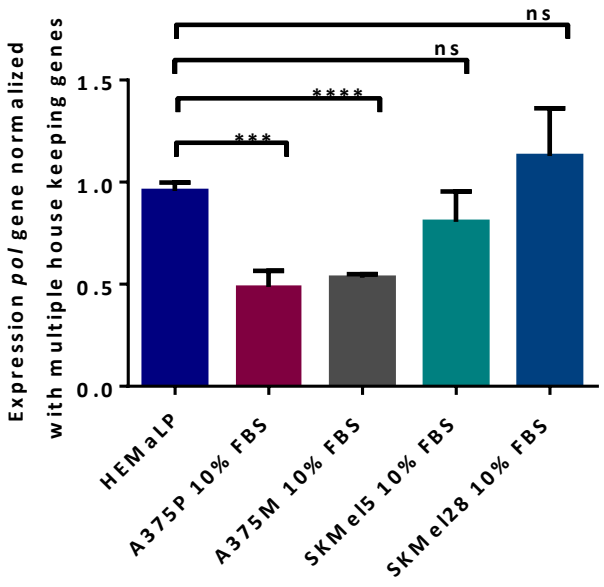


Figure 3.13: Quantitative HERVK *pol* gene expression in melanoma cells and melanocytes.

The *pol* gene was expressed more in melanocytes than in A375P and A375M. At the same time SKMel5 and SKMel28 showed non-significant differences in expression of *pol* in melanocytes. The bars represent relative expression of the target gene (*pol*) normalized with multiple housekeeping genes and using the Pfaffl method. The values plotted are means ± SEMs from three biological repeats. Non-significant (ns), * p <0.05, ** p <0.01, *** p <0.001, **** p <0.0001 vs Control (Student's unpaired t test).

3.3 Discussion

HERVs have been linked extensively with different types of cancer; melanoma (Buscher *et al.*, 2005), breast (Wang-Johanning *et al.*, 2014), prostate (Goering *et al.*, 2015) and colon cancer (Lee *et al.*, 2014a). However, the correlation between HERV-K expression (mRNA and protein) and cancer cells or some other etiological agent playing an important role in cancer progression is still not clearly understood (Huang *et al.*, 2013).

According to (Serafino *et al.*, 2009), the change of phenotype in melanoma cells from adherent to non-adherent (suspension) was readily observed in serum starvation conditions (1% FBS), and can be reduced by inhibition of HERV-K expression.

In this chapter, colony formation and viability assays were carried out to demonstrate adoption of suspension phenotype by the SKMel5 melanoma cell line. Melanoma cells and other cancer cells lines were used to investigate HERV-K *pol* gene expression in two culture conditions, 10% FBS and 1% FBS. We also determined the expression of HERV-K in all the melanoma cells line analysed here and compared this with normal melanocytes.

3.3.1 Validation of cancer cell line identity

Genotyping using polymorphic short tandem repeat (STR) loci and sequence analysis of specific TERT mutations were used to establish the identity of the cell lines studied, as well as provide a genetic means to distinguish the clonal sublines A375P and A375M. Interestingly (Figure3.1), A375P has lost one allele when compared with the ATCC STR genotype data on their website (https://www.lgcstandards-atcc.org/ Quality/ Authentication/Testing/STR%20Analysis), while A375M had both the expected alleles.

This may be related to loss of heterozygosity at the TPOX locus, or perhaps loss of a chromosome arm, during growth in culture. Our analyses could not distinguish these possibilities, but karyotyping may be advisable in future. Nevertheless, this difference does fortuitously enable genetic discrimination between the two clonal derivatives, which is useful.

Similarly, with regards to mutations in the TERT promoter, the TERT promoter was examined for telomerase activating mutations that are commonly seen in melanoma cell lines, particularly at the C228T and C250T positions. Our data showed that the mutation C228T was shared between the two cell lines, while A375P and A375M lack the mutation at C250T. These findings are consistent with Huang *et al.* (2013) who found that the mutation of the TERT promoter at C228T position was recorded in 39% of melanoma cases, while the mutation at C250T position was only recorded in 33%. Thus, in spite of fact that A375P and A375M were derived from the same population of cells (P indicates "progenitor" whereas M indicates the derivative of this cell line), the pattern of mutation at TERT is not a useful diagnostic to distinguish them.

With regards to the identities of the cell lines used in this study, the allelic loss at TPOX in A375 was the only difference between the cell lines and the established STR genotypes, validating that the cell lines used were identical to those available in the ATCC repository.

3.3.2 Phenotype shifts in melanoma and other cancer cells in two conditions

Melanoma cell lines (A375P, A375M, SKMel5 and SKMel28) along with prostate cancer (LNCAP), colon cancer (SW480) and breast cancer (T47D) cell lines were grown under 10% FBS normally, while growth in 1% FBS represented a stress condition.

After 4 days culturing, we found only SKMel5 cells changed their phenotype under 1% FBS: clusters of round-shaped cells detached from the adherent monolayer (Figure3.5), and these cells were able to seed new cultures. However, the rest of the melanoma and other cancer cell lines did not change their morphology except for A375 (P and M) which showed small cluster of cells in suspension.

However, the possibility remained that the cells observed in suspension were simply dead or dying. Thus, colony formation and viability assays were carries out to address these concerns. Our results identified that all cancer cell lines analysed showed significantly reduced numbers of adherent colonies under 1% FBS compared with those formed in 10% FBS. It is clear that low serum conditions impact on colony proliferation perhaps due to limitation of the factors found in serum such as Lactate-dehydrogenase, Cholesterol, vitamins and hormones which support cell growth and adhesion (Masters & Stacey, 2007) .

Recently, some studies have demonstrated that stress conditions can play a role in biological behaviour and growth of melanoma cells (Nadeau *et al.*, 2015; Serafino *et al.*, 2009). In this study, breast cancer cells (T47D) was grown in DMEM media with only 1% fetal calf serum (FBS). The number of cells were reduced compared to 10% FBS as normal condition.

However, the number of T47D cells were significantly decreased when cells treated with Fatty acid synthase (FASN) inhibiter (C75) under 1% FCS culture, compared to cells which were grown under normal conditions (Calvani *et al.*, 2015; Schug *et al.*, 2015; Goering *et al.*, 2015).

Our data show that serum stress conditions change the strength of attachment in cell-cell and cell-surface interaction. A surprising result was that low serum conditions reduced cell surface attachment to such a degree in A375P and A375M cells that they were removed by the fixation process used in colony formation assays, despite the fact that colonies had formed (Figure3.6).

A possible explanation for this might be that under starvation conditions the cells changed their cell surface adhesion proteins; it is possible that cell surface integrin receptors and other surface proteins distribution is altered under 1% FBS. Indeed, (Liang & Dong, 2008) noted that tetherin, integrin and adhesins, which are essential factors in interaction cell-cell and cell-surface adhesion, were reduced in function under shear flow conditions.

In conclusion, it is difficult to determine whether the cellular response is due to 1% FBS or other factors related to the culture environment, but given the established role of cellular stress in altering attachment, and the fact that culture conditions were identical except for the reduction in serum this seems to be a likely explanation.

Only one of the cell lines analysed here (SKMel5) showed a change in phenotype under low serum conditions. While colony-forming assays are appropriate to assess viability in adherent cells, they are not appropriate for suspension cells (or even weakly attached cells as seen for A375 lines in 1% FBS). In addition, it is possible that the population of "suspension cells" observed in SKMel5 cultures are simply dead or dying cells.

To verify that SKMel5 suspension cells were viable under 1% FBS, a cellular viability assay was used (Figure3.7). Our data from the viability assay revealed that SKMel5 cells showing a suspension morphology under 1% FBS, were viable and showed significantly more growth more than cells showing adherent morphology under the same conditions, or the 10% FBS control. In contrast A375P cells showing a suspension phenotype were established to be dead / non-viable cells.

These results are in agreement with the findings of Serafino *et al.* (2009) which showed the melanoma cell line TVM-A12 adopted a suspension phenotype in 1% FBS and showed increased number and size of colonies compared to 10% FBS.

Thus, our data for SKMel5 are similar to those observed for TVM_A12, and strongly suggest that the SKMel5 cell line found a means to proliferate in serum starvation conditions. This result also shows that phenotypic change, under serum starvation is not restricted to the TVM_A12 clonal cell line, and may be frequent for melanoma cell lines.

3.3.3 Quantification of HERV-K *pol* gene expression in melanoma cell lines and other cancer cells, under two conditions

In general, we observed HERV-K *pol* gene expression in all melanoma cells and all the other cancer cell lines examined, in both culture conditions. This suggests that the presence of HERV-K pol transcripts is not a specific biomarker for melanoma.

There is little published data on *pol* expression in melanoma and other cancer cells.

One published study examined the role of *pol* expression in primary melanomas, lymph node metastases, and cutaneous metastases, which were taken from melanoma patients (Muster *et al.*, 2003).

They found a high copy number of HERV-K *pol* transcripts in these cells compared to benign nevus and lymph node cells in healthy individuals and they also found *pol* transcripts in the cytoplasm of 60-90% of melanoma cells, using *in situ* hybridization (ISH).

Our experiments confirmed the expression of the HERV-K *pol* gene in melanoma cell lines, but extended this to other cancer cell lines using qPCR in different culture conditions. This analysis revealed that the modification of phenotype from adherent to suspension morphology under 1% FBS serum, seen in SKMel5 was not associated with increased HERV-K *pol* expression.

Most studies demonstrated expression of HERV-K *env* and *gag* genes, but did not include the *pol* gene or look at a range of other cancer cell lines, (breast, prostate and colon cancer) (Muster *et al.*, 2003; Rhyu *et al.*, 2014; Wang-Johanning *et al.*, 2014; Goering *et al.*, 2015; Lee *et al.*, 2014).

In particular, A375P (progenitor cells) increased *pol* expression under 1% FBS compared with 10% FBS by about 5.6 fold, while no statistically significant difference in expression between the conditions was observed in A375M (metastatic cells). Surprisingly, although these cells are almost genetically identical, they expressed different levels of *pol* in starvation conditions. This result may be explained by the fact that serum starvation may impact A375P more than A375M despite their common origin and genetic similarity. Huang *et al.* (2013) mentioned that melanoma cells can gain survival and growth advantages in stressful cellular environments, which can subsequently contribute to metastasis, resulting in much poorer patient prognosis.

Another possible explanation for our results is that an epigenetic state may be changed under 1%FBS. A recent study observed that hypomethylation plays an important role in increased expression HERV-K mRNA when the DNA methyltransferase enzymes (DNMT) was knocked out (Stengel *et al.,* 2010). On other hand, SKMel5 and SKMel28 showed reduced expression of *pol* compared with 10%FBS culture (Figure3.6). It is possible, therefore, that tumour suppression genes such as TP53 may contribute to the reduction of HERV-K *pol* gene expression in these cells.

Taken together our results above indicate that the expression of HERV-K is not necessarily up regulated under starvation conditions that induce modification of cellular phenotype from adherent to suspension which is not specifically associated with increased or reduced expression, in our hands.

This finding contrasts with previously published work on HERV-K expression in melanoma cells (TVM-A12 and M14) under starvation conditions. (Serafino *et al.*, 2009) observed HERV-K overexpression, when melanoma cells transition from adherent to suspension phenotype, in 1% FBS. However, there are several possible reasons for the conflict. Firstly, the Serafino study used patient derived cell lines, which were established in their lab, and measured *env* gene expression in these cells. Thus, the extrapolation that HERV-K expression drives morphological changes in all melanomas is not justified by this data. On the other hand, all the other types of cancer cell lines that we analysed under 1% FBS showed increased expression of *pol* and SW480 showed a statistically significant increased expression of *pol* (Figure3.10*)*.

Many studies have demonstrated the expression of HERV-K by investigating the *gag* and *env* genes, but did not include the *pol* gene. For example, in breast cancer cells HERV-K *env* mRNA was highly expressed in patients who have advanced breast cancer

compared to patients who have early stages of breast cancer (Wang-Johanning *et al.*, 2014). Moreover, Rhyu *et al.* (2014) reported that *env* gene expression was increased in patients who were treated with radiotherapy compared with chemotherapy. It may be that increased expression is related to a change in environmental factors such as development of cancer or types of drugs.

In prostate cancer, DNA methylation and steroid hormones can play important roles to induce HERV-K gag expression in LNCAP cells. Expression of HERV-K *gag* in patients who have prostate cancer was investigated with regards to a specific locus, ch22q11.3. Hypomethylation and steroid hormones together induced high levels of HERV-K *gag* protein expression in these cancer cells compared to normal cells. It has been suggested, as a result of these studies, that HERV-K *gag* expression in ch22q11.3 can be used as a good biomarker for diagnostic prostate cancer (Reis *et al.*, 2013; Goering *et al.*, 2015).

Taken together all the published papers concerning breast and prostate cancer noted that HERV-K expression can be used as a biomarker as increase expression is associated with abnormal cells. Our data suggest this is not the case for all melanoma cell lines and may also be a cell line specific characteristic in other cancers as well.

With regards to SKMel5 (Figure3.12), significant difference in expression of *pol* was observed in1% FBS, when the cells adopted a suspension phenotype but there was a downwards trend, rather than an increase.

It could be possible, the change in phenotype may be related to mutation or deletion in some transcription factors sites such as Sp1 and Sp3 and reduced level of expression (Fuchs *et al*., 2011). Also changes in the methylation status of the HERV-K LTR under 1%FBS may decrease the initiation of transcription.

3.3.4 Investigation of HERV-K *pol* expression in melanoma cells and melanocytes (HEMaLP)

Our results showed that the absolute level of *pol* expression in A375P and A375M was significantly reduced compared with melanocytes. SKMel5 and SKMel28 by contrast showed similar levels of *pol* expression to melanocytes (Figure3.13).

These results imply that there was no difference in the level of HERV-K mRNA between melanoma cells and normal melanocytes, raising the possibility that HERV-K expression is not related with the stage of development of melanocytes in melanoma

In contrast, Muster *et al.* (2003) demonstrated that HERV-K *pol* sequences were present in primary and cutaneous metastases by *in situ* hybridization (ISH), while healthy individuals (benign nevus) showed negative results.

In addition, Buscher *et al.* (2005) reported that HERV-K *env* gene expression was increased in 45% of metastatic melanoma and 44% of melanoma cells, compared to human Epidermal Melanocytes, neonatal, lightly pigmented (HEMn-LP) donor, However, (HEMn-LP) *rec* was detected, while the *env* gene was not investigated (Buscher *et al.*, 2005). In 2006 Buscher *et al*., found *np9* expression in 29% of metastatic melanoma and 21% of melanoma cells using RT-PCR.(Buscher *et al.*, 2006).

In addition to melanoma, HERV-K *pol* is expressed in breast, prostate and colon cancer. Furthermore, the response to starvation conditions is not simply related to increased expression of HERV-K and changes in phenotype are limited to only particular melanoma cell lines, such as SKMel5.

However, here we only studied the transcripts from one HERV-K gene in melanoma and other cancer cells, the effects of HERV-K protein expression are explored in detail in the next chapter.

4 Analysis of HERV-K protein expression in cancer cells

4.1 Introduction

Although, the HERV-K (HML-2) family is exceptional because intact proviruses encode proteins that have been associated with different types of cancer, the principle of HERV-K (HML-2) transcription remains poorly understood (Lavie *et al.*, 2005; Agoni *et al.*, 2013; Goering *et al.*, 2015; Argaw-Denboba *et al.*, 2017).

In recent years, there has been increasing interest in melanoma because it is a common skin cancer, patients who have metastatic melanoma are difficult to treat, and consequently their survival is low (Riverso *et al.*, 2017). It has been postulated that expression of HERV-K proteins may contribute to malignant transformation in melanoma cells, through the action of HERV proteins altering these cells' morphology and behaviour (Serafino *et al.*, 2009).

Other groups have built on this work by defining exactly which genes are responsible for the HML-2 signal seen in different types of cancer. The expression of the HERV *pol* gene, and the *gag*, *rec* and *env* proteins has been demonstrated in primary and metastatic melanoma tissue (Muster *et al.*, 2003).

When HERV-K gene expression was investigated directly the expression of spliced *env* and *rec* was found in 45% of melanoma metastases, and in 44% of melanoma cell lines In contrast, normal human epidermal melanocytes of light and dark pigmentation (HEMn-LP, HEMn-DP) did not express the HERV-K *env* mRNA but *rec* mRNA, was detected in both of the darkly pigmented and the lightly pigment cell lines. Moreover, proteins encoded by the *env*, *rec* and *np9* genes were found in metastatic melanomas and melanoma cell lines by immunohistochemistry IHC (Buscher *et al*., 2005).

Another study found that the HERV-K *env* protein was expressed in melanoma cells, but also that, *rec* and *np9* were also detected in normal melanocytes as well as melanoma cells (Buscher *et al.*, 2005; Buscher *et al.*, 2006).

In addition, while UV exposure is a well-established risk factor for melanoma, in fact the level of transcription of HERV-K *rec* and *np9* genes in normal human melanocytes cells increased expression after UV exposure. However, in melanoma cells, HERV-K *rec* and *np9* reduced or did not-changed expression when cells are irradiated with UV light (Reiche *et al*., 2010).

Schmitt *et al.* (2013) found that the ERV6 (Endogenous Retrovirus Group K Member 6) in 7p22.1 was cloned at high frequency in melanoma cDNA libraries, ranging from 21% to 75% in seven of the melanoma samples, but did not observe transcription in melanocyte cell lines. Other proviral loci such as ERVK-1, ERVK-7, and ERVK-5 were cloned at frequencies ranging from 29% to 53% from the analysed above two melanoma samples. This observation suggests that the high relative cDNA cloning frequencies (up to 92%), is associated with higher level of transcription of those particular loci. It is possible concluded that, the ERVK-6 in chromosome 7 was responsible for the majority of the encoded *gag* and *env* proteins in melanoma cells, but was silenced in melanocytes, as the *gag* and *env* proteins were not detected in melanocytes (Schmitt *et al.*, 2013).

Recently, there has been renewed interest in the *env* protein. Krishnamurthy *et al.* (2015) demonstrated that HERV-K *env* protein was could be good target for the T-cell mediated treatment of melanoma. In this study, the authors used T cells which were genetically engineered to express a chimeric antigen receptor (CAR+ T), to bind ERV *env* protein expressed on the surface on melanoma cells, and direct their lysis in an *env*-dependent manner. While this study was carried out in a mouse model, the target cells were a melanoma cell line, and significant tumour regression was achieved. Thus, if it is feasible to apply this approach in humans, that this mechanism could be used to reduce melanoma development in clinical contexts.

On other hand, it has been noted that MAPK (RAS/RAF/MK/ERK) signalling pathways play a vital role in cell proliferation (Nazarian *et al.*, 2010). However, 60% of melanomas harbour BRAF mutations (typically V600E), which leads to constitutive activation of ERK signal pathway (Nazarian *et al.*, 2010; Tran *et al.*, 2015).

BRAF (V600E) inhibitors can be used as pharmacological agents, however, these inhibitors were shown to be resistant to immunotherapy (Riverso *et al.*, 2017). Recently, MEK/ERK inhibitors have been used as alternative way to block activation of mutated BRAF/MEK/pERK in melanoma (Tran *et al.*, 2015). Also, a recent study found that the expression of HERV-K correlated with pERK activation in melanoma cells and the combination of pERK and CDK4 inhibitors reduced HERV-K *env* protein expression. It could suggest that targeting HERV-K, pMEK and CDK4, may produce more effective therapeutic effects than using a single target.(Li *et al.*, 2010). This mechanism is outlined in Figure 4.1.

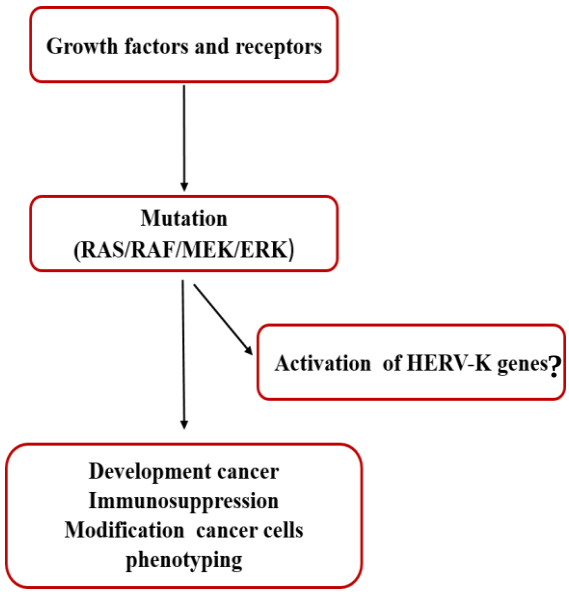


Figure 4.1: Relationship between HERV K expression and signaling pathways.
Mutations in the BRAF→MEK→ERK mitogen cause disruptions in MAPK pathway and constitutive activation of the BRAF kinase, results in uncontrolled cell proliferation. The same time, HERV-K protein expression correlated with pERK activation, suggesting HERV expression may regulate growth downstream of pERK.

In experiments reported in this thesis, HERV-K mRNA was detected in melanoma cells and other cancer cells and the level of expression of the *pol* gene was modulated by growth under conditions of serum restriction, although not in the way expected from the literature.

However, while HERV-K proviruses express their proteins from a multi-cistronic mRNA it possible that *pol* expression is not a reliable proxy for HERV-K gene expression. To address this expression levels of HERV-K proteins were measured in 10%FBS and 1%FBS culture condition, to investigate their correlation with mRNA levels. In addition, we attempted to discover the correlation between HERV-K *gag* protein expression and the pERK signalling pathway in melanoma cells to investigate the hypothesised relationship between HERV-K expression and aberrant intracellular signalling.

4.2 Results

4.2.1 Comparing HERV-K *gag* protein levels in melanoma and other cancer cells under two culture conditions

To determine whether HERV K specific proteins were detectable in melanoma cell line and other cancer cell line, we performed western blot analysis with antibodies specific for HERV-K proteins. A375P, A375M, SKMel5 and SKMel28 (melanoma cells) and LNCAP, SW480 and T47D (other types of cancer cell lines) were cultured in 10% FBS and 1% FBS, to establish whether the changes in HERV-K mRNA expression were related to protein expression. Lysates were isolated using Laemelli buffer and processed for western blot analysis. Membranes were stained using a *gag* monoclonal antibody from AUSTRAL Biologicals Company to detect HERV K *gag* and capsid antigens. The western blot analysis showed the expression of a *gag* antibody binding protein at the appropriate mass (82kDa), in all the cancer cell lines analysed, with the exception of T47D. There no other *gag* antibody binding protein signals in the lane. Western blots were exposed for a maximum of three minutes. Also, generally *gag* expression was higher under 10% FBS, particularly in SKMel28, compare to 1% FBS condition, as illustrated in Figure 4.2. By contrast no *gag* signal was detected in T47D, in either condition.

We performed western blot analysis to quantify the protein levels of *gag* and found that most cancer cell lines, whether they were melanomas or not, significantly decreased *gag* expression in 1% FBS culture, compared with the 10% FBS culture. The prostate cancer cell line LNCAP showed a weak signal, but there was no significant difference in *gag* expression between the two conditions (Figure 4.3). All these analyses used tubulin as a normalisation control.

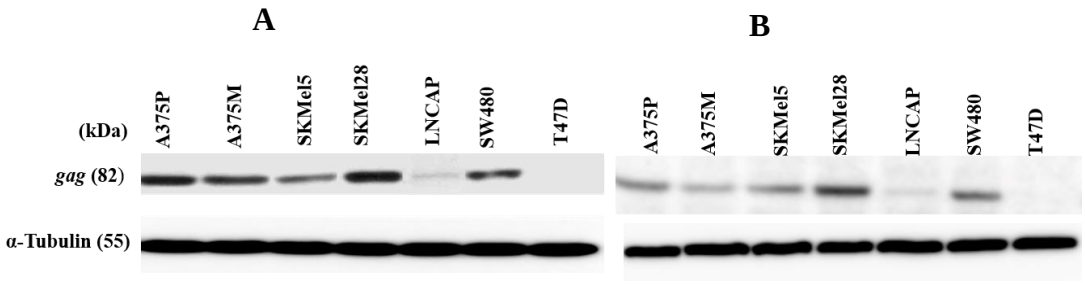


Figure 4.2: Detection of HERV K *gag* protein in two culture conditions.

(A) Melanoma and other cancer cell lines were grown with 10% FBS in the culture medium. Cells were lysed and 20 µg of proteins analysed by western blotting. Blots were probed with antibody against *gag*. α- Tubulin was used as the protein loading control. The expression of *gag* protein was detected in all melanoma cells. LNCAP showed a weak signal compared with SW480 example. **(B)** The same cancer cells were grown as adherent layers under 1% FBS. The expression of *gag* protein was down-regulated in all melanoma and cancer cell lines. T47D showed no detectable expression of *gag* protein in both conditions. Representative images of multiple replicate experiments (n = 3) are shown.

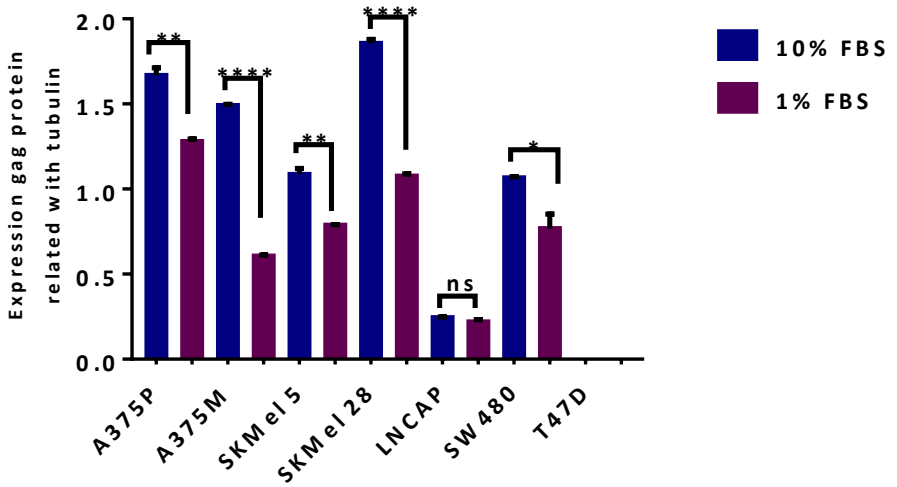


Figure 4.3: Quantification *gag* proteins in two culture conditions.

The bar chart shows *gag* protein levels, quantified using the ImageJ software from the western blot results. The values plotted are means ± SEMs. Statistical analysis was by Student's unpaired t test. Non-significant (ns),* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ vs Control.

Taken together these data suggest HERVK *gag* protein expression is not up regulated in serum starvation conditions, in fact being significantly reduced in all melanoma cell lines, by this treatment. However, we did observe consistently higher *gag* protein expression in melanoma cells than in other cancer cells, with the exception of the colon cancer cell line SW480. ANOVA of the *gag* protein expression level showed that under 10% FBS all melanoma cell lines showed significantly higher expression than SW480, with the exception of SKMel5 (Figure 4.4, One way A NOVA of all cell lines under 10% FBS)

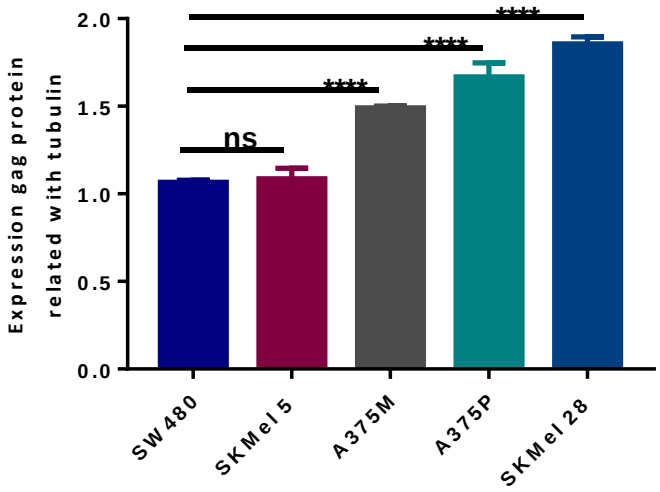


Figure 4.4: Quantification *gag* proteins in SW480 and melanoma cells under 10%FBS.

The bar chart shows *gag* protein levels in SW480 and melanoma cells under normal condition, quantified using the ImageJ software from the western blot results. The values plotted are means ± SEMs. Statistical analysis was by one way ANOVA. Non-significant (ns),* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ vs Control.

4.2.2 Comparing HERV-K *env* protein levels in melanoma and other cancer cells under two culture conditions

While the previous experiments used the *gag* protein, as a proxy for HERV-K provirus expression, most analyses of altered HERV-K expression in melanoma have focussed on the *env* protein. Although *env* is on a separate spliced form of the viral transcript – bias in alternative splicing, in favour of the *env* form could produce differences translated from the same proviral transcript, they do share a promoter and 5'UTR so we would expect their protein levels to correlate to some degree. As a result, we considered whether *env* protein expression changed in cancer cell lines, including melanomas, in normal and serum starvation conditions. To address this question, we carried out western blotting with an *env* specific monoclonal antibody from AUSTRAL Biologicals Company. No other *env* antibody binding protein signals were detected in the lane. Western blots were exposed for a maximum of three minutes.

Western blot analysis showed that, surprisingly T47D showed robust *env* protein expression, which increased under serum starvation (Figure 4.5) By contrast all melanoma cells and LNCAP expressed *env* protein at low levels, but for SW480 (colon cancer) we did not detect any signal from the *env* protein (Figure 4.4 A).

Next, we sought to investigate the HERV K *env* protein in cancer cells, but growing in culture medium containing 1%FBS, as adherent layers. We used lysates from cells that were grown under 1%FBS. The same assay was carried out to observe any changes of *env* protein expression under these conditions. Western blot analysis revealed that SKMel5 up regulated expression of the *env* protein in 1% FBS, when compared with SKMel5 under normal (10% FBS) conditions (Figure 4.5B). Moreover, T47D showed

increased expression of *env* under 1% FBS, compared to 10% FBS culture. The rest of the cell lines, including LNCAP showed down regulation of *env*.

Quantification of *env* protein levels found that prostate cancer, colon cancer and melanoma cell lines A375M and SKMel28 showed decreased expression (a statistically significant reduction in the case of LNCAP, SW480, A375M, and SkMel28) of *env* protein in 1%FBS culture, compared with the 10%FBS culture (Figure 4.6). In conclusion, the present results demonstrated that the HERV K *env* protein in SKMel5 under 1%FBS condition showed a significant increase in concentration, which is consistent with the increase in *env* expression that Serafino et al described in their paper, and was associated with a change in cellular phenotype. Interestingly T47D, a breast cancer cell line, showed divergent expression of HERV-K proteins (being *gag* negative and *env* positive, by western blot) despite these being alternatively spliced versions of the same proviral pre-mRNA, driven by the same promoter, at least according to established models of HERV-K replication.

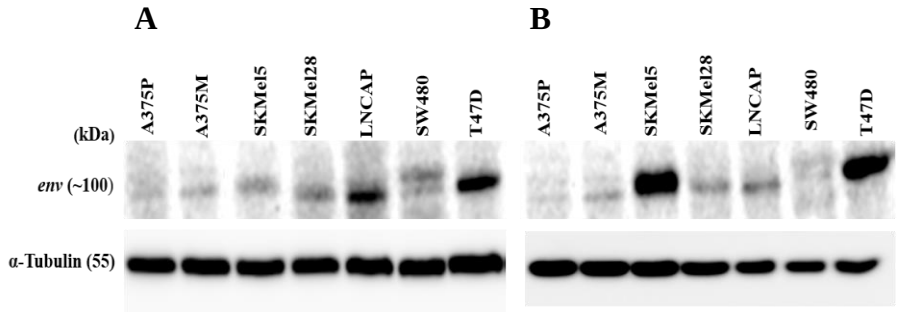


Figure 4.5: Detection HERV K *env* protein in two conditions.

(A) Melanoma and other cancer cells were grown with 10% FBS in the culture medium. Cells were lysed and 20 µg of proteins expression was detected via western blotting. Blots were probed with antibody against *env* α- Tubulin was used as protein loading control. The expression of env protein was detected in T47D, while melanoma cells showed a weak signal compared with T47D. **(B)** The same cells were grown 1% FBS in the culture medium. The data showed that SKMel5 displayed higher *env* protein expression than other melanoma cell lines and T47D. Also, LNCAP displayed reduced *env* protein expression compared with the levels observed in 10%FBS. Representative images of multiple replicate experiments (n = 3) are shown.

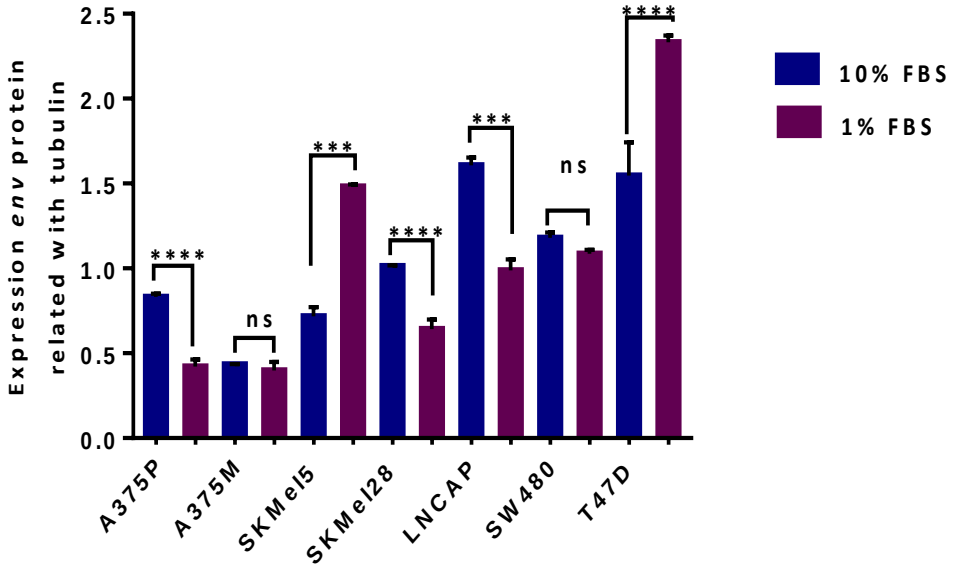


Figure 4.6: Quantification level of *env* proteins in two conditions.

The bar chart shows the *env* protein levels quantified using ImageJ software, according to the western blot results. The values plotted are means ± SEMs compared by Student's unpaired t test. Non-significant (ns),* p <0.05, ** p <0.01, *** p <0.001, **** p <0.0001vs Control.

4.2.3 Expression of protein levels (*gag* and *env*) in SKMel5 suspension cells under 1%FBS

Having detected a significant change in HERV-K protein levels in adherent cells under serum starvation (1% FBS), we next wanted to investigate HERV-K *gag* and *env* proteins in the cells of SKMel5, showing their modified suspension phenotype. It was of particular interest to find out whether SKMel5 cells, when expressing the suspension phenotype, increased expression of *gag* and / or *env* proteins. Therefore, SKMel5cells in 10% FBS and in 1% FBS showing adherent and suspension phenotypes were separated and queried with *env* and *gag* antibodies. In addition, these lysates were transferred to one membrane to ensure identical conditions and exposure time. We observed that SKMel5 cell has higher HERV-K *env* protein expression in 1% FBS as an adherent layer as compared to 10% FBS (Figure 4.7 A).

However, when analysing the suspended cells from the same 1% FBS cultures we did not detect any *env* protein signal. In addition, as can be seen in figure (4.7 B), *gag* protein expression is lower in SKMel5 adherent cells in 1% FBS, compared to 10%, and lower still in suspension cells. Quantifying these results, using α-tubulin as a loading control, showed that the observed changes were statistically significant by one way ANOVA (Figure 4.8 A and B).

Interestingly, the *gag* protein expression results parallel the *pol* qPCR results reported in Chapter 3, but contrast with changes in *env* expression. This finding makes it clear that gene expression (*pol*) and protein (*gag*) levels of two of the three major HERV-K genes were not related to the observed changes in phenotype under serum starvation-induced stress conditions. In sharp contrast, increased *env* protein expression does correlate with serum starvation-induced stress, consistent with previous reports for some melanoma cell lines. This finding illustrates de-coupling of HERV-K gene expression and protein levels, despite the genes being encoded by the same proviral transcript, and driven by a common promoter (the HERV-K LTR).

One possibility is *env* protein expression is independent of the HERV-K LTR promoter, and that its modulation does not reflect HERV provirus activation. Importantly our results show that even when melanoma cell line cultures alter their phenotype, while the adherent cells show the previously observed upregulation of *env* under 1% FBS, the cells that actually express the altered phenotype have barely detectable *env* protein levels.

This distinction may have been missed in previous analyses, where the adherent and suspension cells were assayed together. Nevertheless, our data are consistent with previous studies showing that HERV-K *env* protein upregulation is correlated with phenotype modulation, at least in some melanoma cell lines. The significance of these findings will be discussed at the end of the chapter.

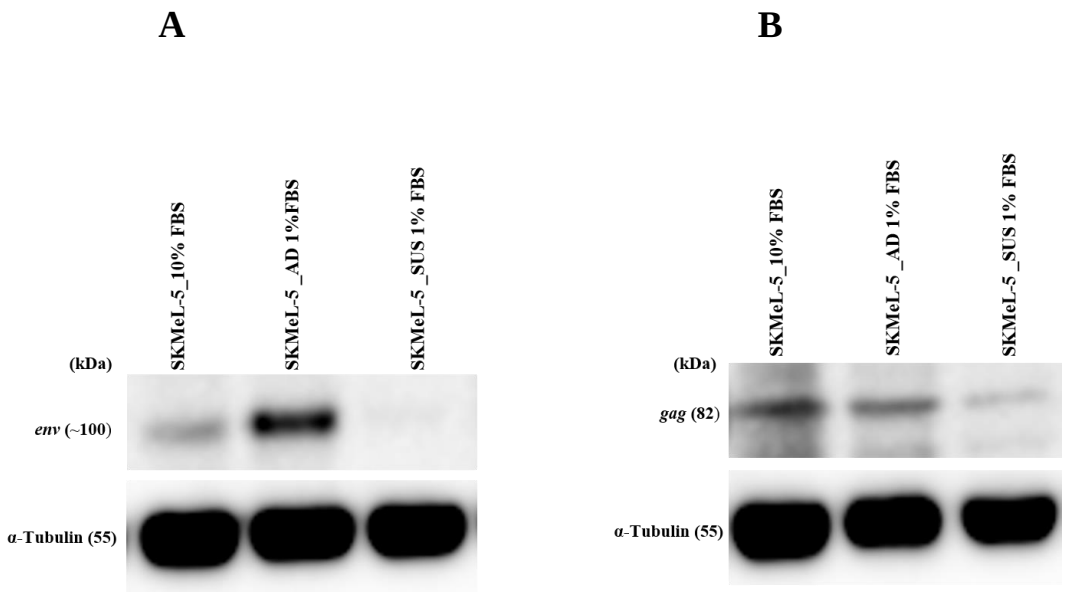


Figure 4.7: Detection HERV- K *env* and *gag* proteins in SKMel5 suspension.

(A) 20 µg of SKMel5 proteins in two conditions was analysed by western blot. Blot was probed with antibody against *env*. α- Tubulin was used as protein loading control. The expression of *env* protein was detected in SKMel5 under 1%FBS (adherent) compare to 10% FBS and 1% FBA (suspension). **(B)** The same cells were used with *gag* antibody. The data showed that SKMel5 reduced expression in 1% FBS in two phenotypes. Representative images of multiple replicate experiments (n = 3) are shown.

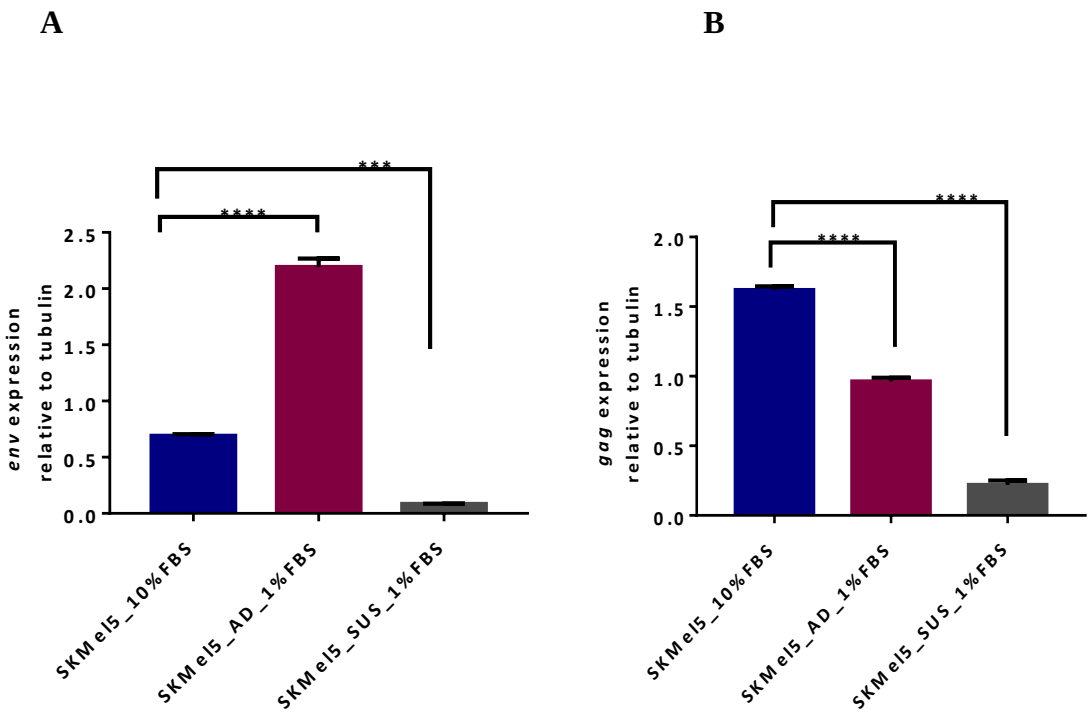


Figure 4.8: Quantification level of SKMel 5 *env* proteins in two conditions

(A)The bar chart shows the *env* protein levels quantified using ImageJ software according to the western blot results. **(B)** Quantification of *gag* protein in SKMel5. Our data showed that the *env* protein was a significantly upregulated in SKMel5 under starvation conditions, in contrast to the *gag* protein, which showed decreased expression, step by step according to the type of condition and phenotype. The values plotted are mean ± SEM, compared by one-way ANOVA. Non-significant (ns),* p <0.05, ** p <0.01, *** p <0.001 vs Control.

To further characterise divergent expression of *gag* and *env* proteins, their cellular localisation was examined using fluorescence microscopy. SKMel5 cells were cultivated under two conditions (10% FBS and 1% FBS) and prepared for staining with fluorescent secondary antibodies to the same *gag* and *env* primary antibodies used for western blots. Due to the need to fix cells to the slide surface through the washing steps only the adherent SKMel5 cells could be analysed.

Any suspended cells in 1% FBS cultures would have been lost. Figure 4.9 A and B show the localisation of *env* (A) and *gag* (B) proteins relative to the DAPI stained nuclei of adherent SKMel5 cells in 10% and 1% FBS. These experiments qualitatively confirmed the previous experiment (4.2.3) and found the *gag* and *env* were localised to both the nucleus and the cytoplasm. It is evident from the representative images presented, that *env*-specific staining is more intense under 1% FBS and *gag*-specific staining more intense under 10% FBS. However, this result contrasts with previous reports (Muster *et al*., 2003) where *gag* and *env* proteins were only found in the cytoplasm.

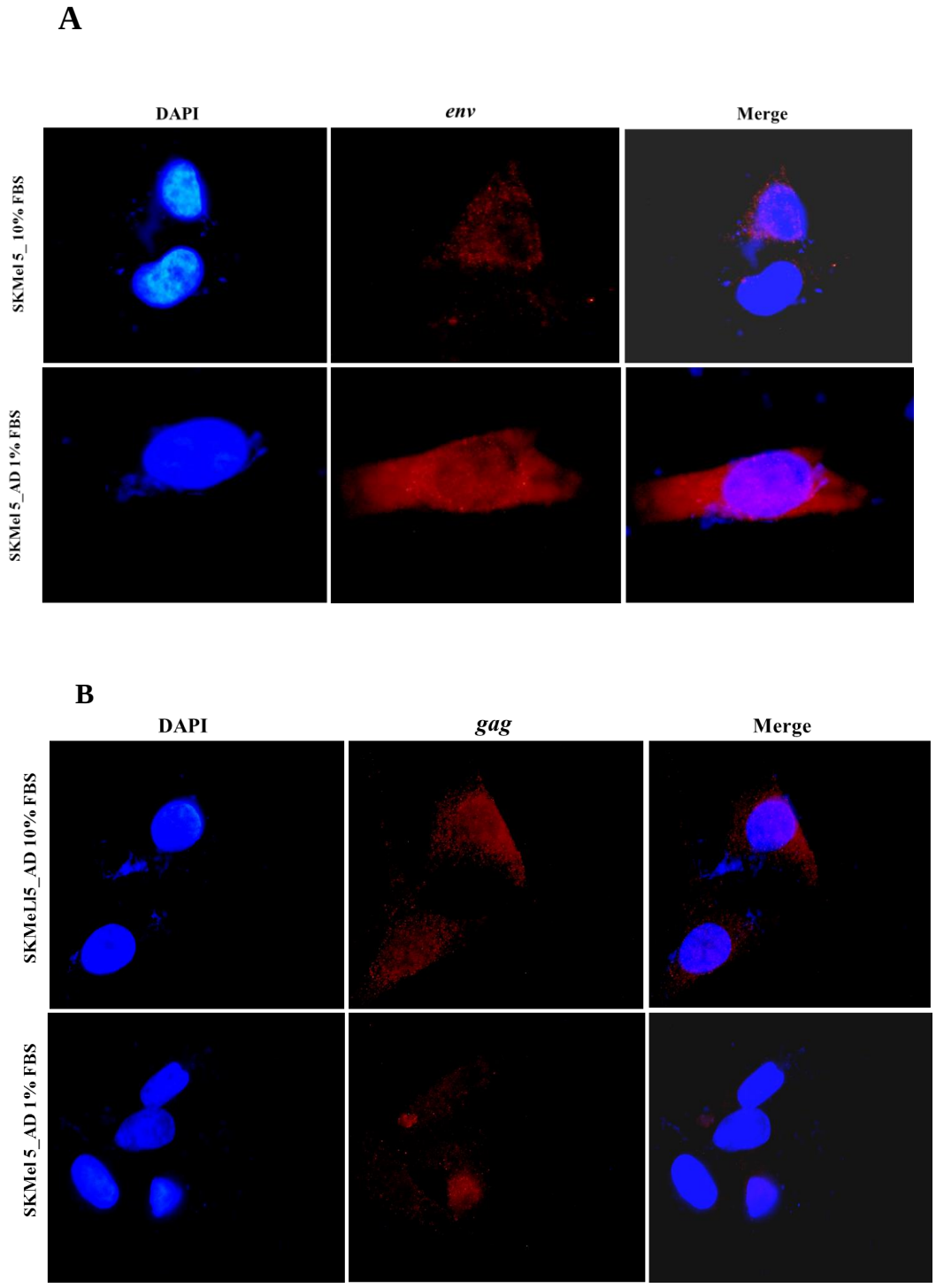


Figure 4.9: Localisation of *gag* and *env* in SKMel5 in two conditions.

SKMel5 cells were maintained in the presence of 10% FBS and 1% FBS for 72 hours, fixed, permeabilised, and stained with HERV-K specific antibodies against *gag* and *env* proteins and DAPI (to visualise nuclei) using immunofluorescence. Red = *gag* or *env*; Blue = DAPI. The capture time was 83millisecond (ms).

4.2.4 Protein expression levels of HERV K *gag* and *env* in primary melanocytes

Our previous experiments indicated HERV K genes and proteins in melanoma cell lines showed up or down regulation depending on conditions and cellular phenotype, generally in agreement with previous published reports. However, these changes were far from consistent across melanoma cell lines, and comparison of the direction of changes in gene and protein expression suggested de-coupling between the HERV gene expressions. To explore this further we examined relative expression levels of the *gag* and *env* proteins in normal primary melanocytes (HEMaLP) in comparison to melanoma cell lines under normal culture conditions (10% FBS).

Four melanoma cell lines (A375P, A375M, SKMel5 and SKMe28), HEMaLP, with T47D as a control, were analysed by western blotting. As shown in Figure 4.10 *gag* protein expression was readily detectable in both melanoma and melanocytes, while T47D did not show any signal. However, quantification of protein levels of *gag*, normalised to α-Tubulin indicated that melanoma cell lines expressed significantly more *gag* protein more than HEMaLP cell (Figure 4.11). These results demonstrated that the HERV- K *gag* protein is detectable in all melanoma cell lines tested and melanocytes, but expressed at significantly different levels. This suggests that *gag*, like *env,* could be a biomarker for melanoma.

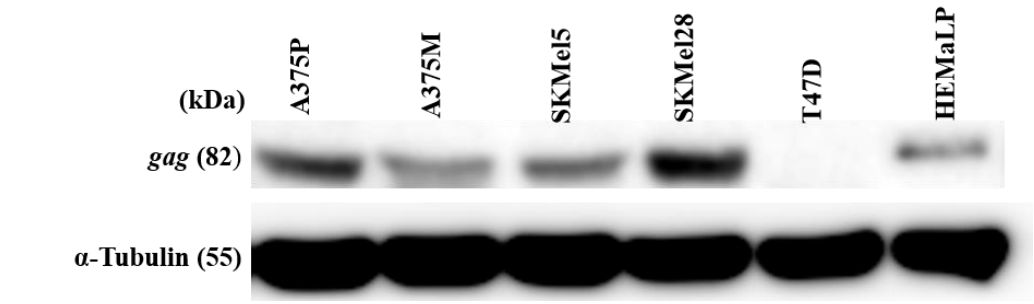


Figure 4.10: Detection HERV K *gag* protein in melanocytes.

Melanoma cell lines and melanocytes were grown in 10% FBS. Cells were lysed and 20 µg of proteins analysed via western blotting. Blots were probed with an antibody against *gag*. α-Tubulin was used as protein loading control. The expression of *gag* protein was detected in all melanoma cell lines and melanocytes. T47D (a breast cancer cell line) was used as a negative control. Representative images of multiple replicate experiments (n = 3) are shown.

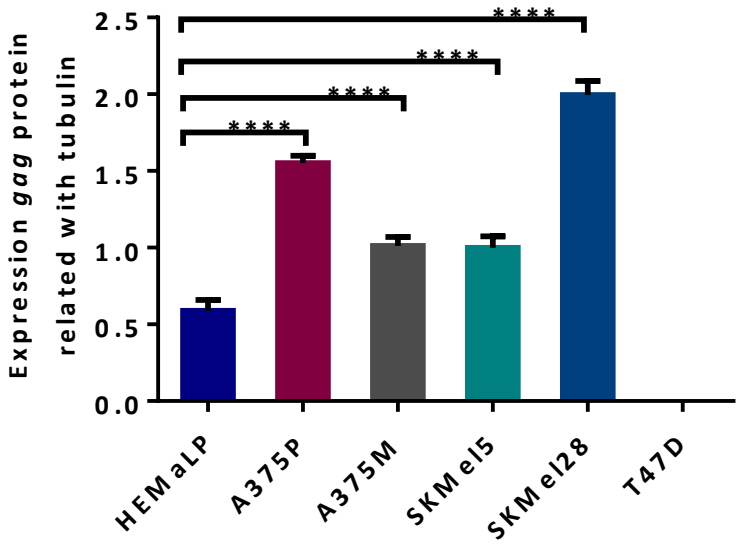


Figure 4.11: Quantification level of melanoma and melanocytes *gag* protein.

The bar chart shows the *gag* protein levels quantified using ImageJ software according to the western blot results. Our data showed that the *gag* protein was highly expressed in melanoma cells, in contrast to normal melanocytes (HEMaLP). The values plotted are means ± SEMs, compared by one-way ANOVA. Non-significant (ns),* p <0.05, ** p <0.01, *** p <0.001, **** p <0.0001 vs Control.

With regards to the *env* protein, the same cell lysates were used to compare *env* protein levels in melanoma cell lines and melanocytes. The western blots showed that all melanoma cell lines expressed *env*, while in melanocytes *env* expression was undetectable (Figure 4.12). These results clearly point to HERV-K *env* expression as a good biomarker for melanoma, as has been previously observed, and perhaps in preference to *gag*, because is the difference in *env* expression level is greater between melanocytes and melanoma. However, on quantification of protein levels of *env* it was observed that A375P showed significantly higher expression of *env* protein, more than other melanoma cells (Figure 4.13). Also, A375M and SKMel5 were much closer in *env* expression level, similar to the result for the *gag* protein.

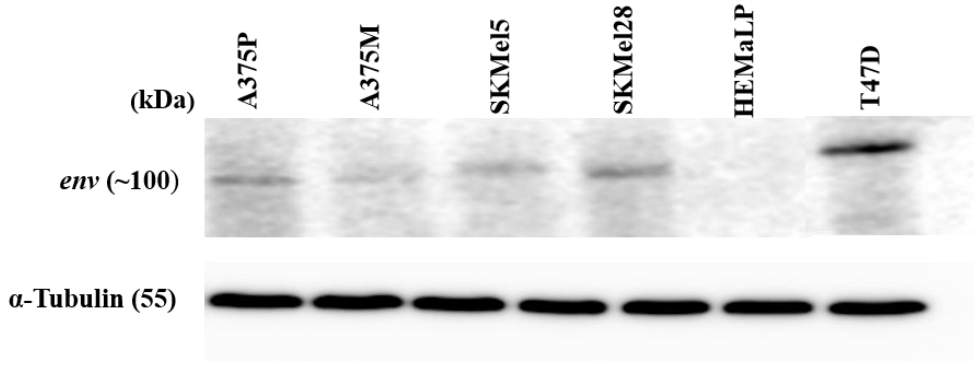


Figure 4.12: Detection HERV K *env* protein in melanocytes.

Melanoma cell lines and melanocytes cells were grown in 10% FBS. Cells were lysed and 20 µg of proteins analysed via western blotting. Blots were probed with an antibody against *env*. α-Tubulin was used as protein loading control. The *env* protein was weakly expressed in all melanoma cell lines, but not in melanocytes. T47D was used as positive control. Representative images of multiple replicate experiments (n = 3) are shown.

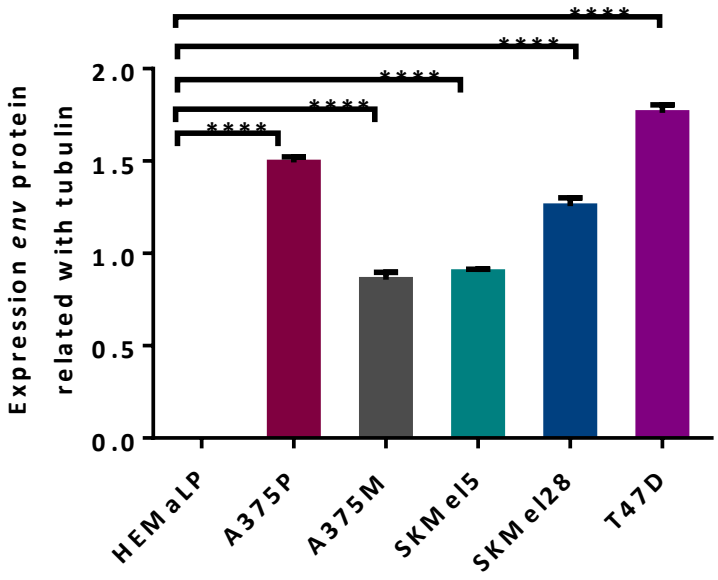
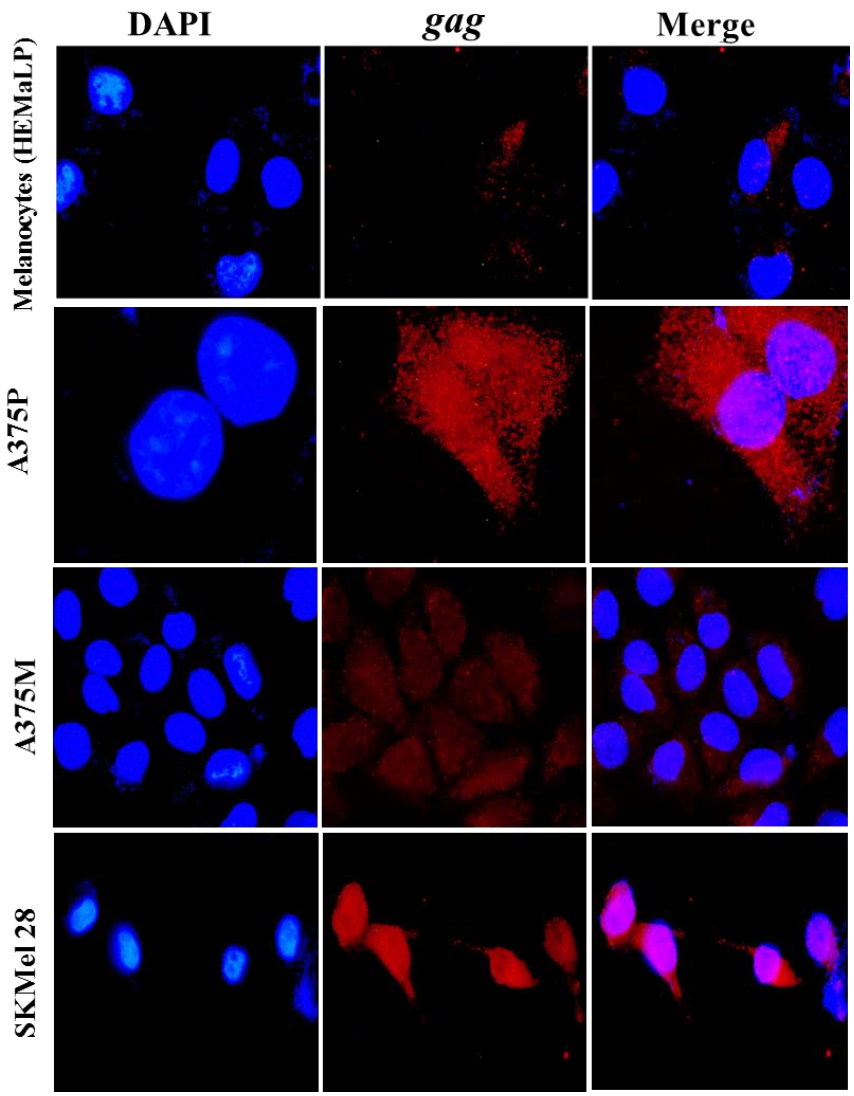


Figure 4.13: Quantification of melanoma cell line and melanocyte *env* protein expression.

The bar shows the *env* protein levels quantified using ImageJ software, according to the western blot results. Our data observed that the *env* protein was highly expressed in melanoma cells, but not detected in melanocytes. The values plotted are means ± SEMs, compared by one-way ANOVA. Non-significant (ns),* p <0.05, ** p <0.01, *** p <0.001, **** p <0.0001vs Control.

To validate the western blotting experiments, immunofluorescent staining for *gag* and *env* antibodies with the A375P, A375M and SKMel28 cell lines and melanocytes was performed. All cells were cultivated in 10% FBS. As shown in Figure 4.14, *gag* (4.14A) and *env* (4.14 B) were localised in the nucleus and cytoplasm.

A



B

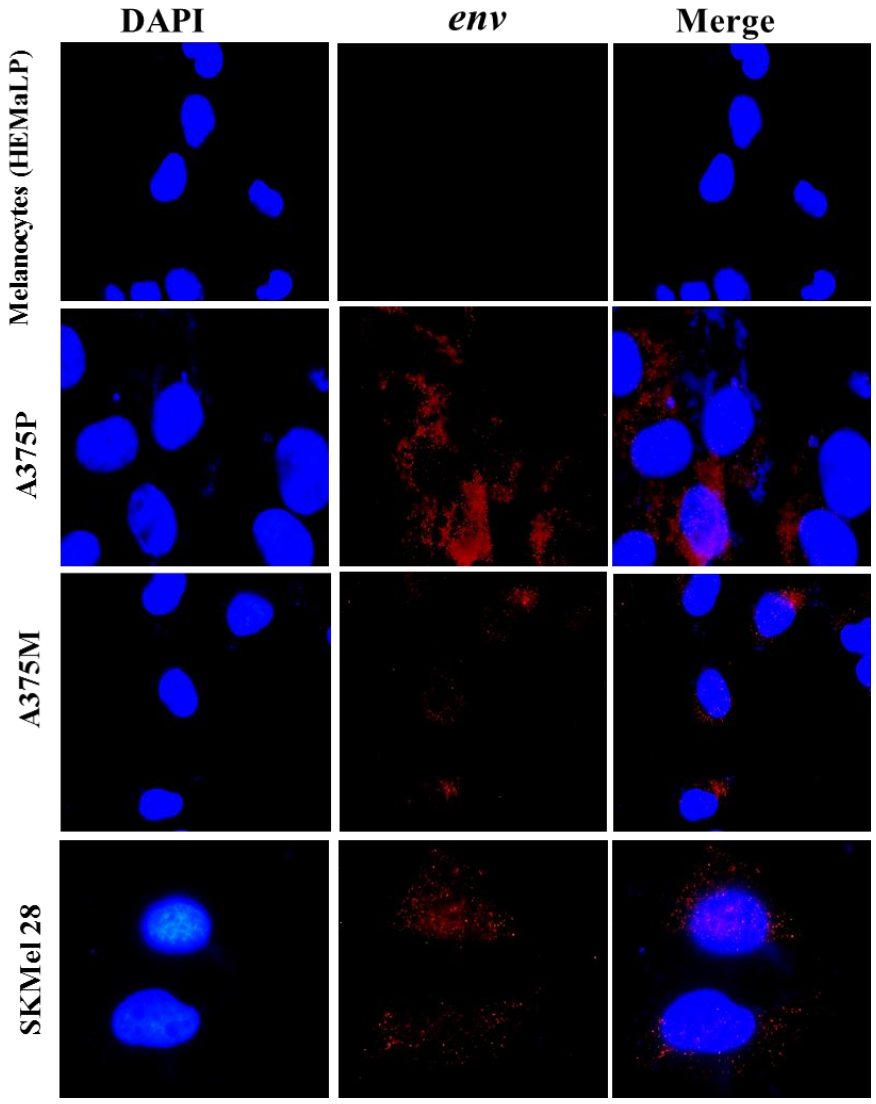


Figure 4.14: Localization of *gag* and *env* in melanoma cell lines and melanocytes.

(A) *gag* antibody with melanoma and melanocytes cells, **(B)** *env* antibody with melanoma cell lines and melanocytes cells . All cells were maintained in the presence of 10% FBS for 72 hours, fixed, permeabilised, and stained with HERV-K specific antibodies against *gag* and *env* and DAPI (to visualise nuclei) using immunofluorescence. Red = *gag* or *env*; Blue = DAPI. The capture time was 83millisecond (ms).

The results presented here demonstrate that serum starvation conditions induced transition to a suspension phenotype, but this seems restricted to a minority of melanoma cell lines, such as SKMel5 which changed its phenotype. However, we did observe that 1% FBS culture did increase HERV-K *env* protein in SKMel5, but only in

those cells with an adherent phenotype. Surprisingly, normal melanocytes were recorded as showing low, but detectable expression of *gag*, while they did not express *env*. It may be the case therefore that these variations of expression of melanocytes can be used HERV-K as a clinical biomarker for melanoma.

4.2.5 Expression of HERV-K proteins in primary and metastatic tissue

HERV-K *gag* and *env* expression were investigated in two stages of melanoma, by using Immunohistochemistry (IHC). Eight patient samples, 4 samples of the benign nevus stage and 4 of metastatic melanoma, were used in his experiment. The first step was optimization of antibody dilutions (1/100, 1/250, 1/500 and 1/1000) to get the best working dilution. Optimisation was necessary as IHC staining can be confounded by natural melanin expression in these cell types. Unfortunately, the *gag* antibody could not be optimised in the time available (data not shown), while *env* antibody gave specific staining at a dilution of 1/250.

Our data showed HERV-K *env* protein was detected in metastatic melanoma tissue (Figure 4.15A and B), compared to nevus stage (Figure 4.16A and B). For further confirmation, the same section of tissue was stained using 1xTBS instead of the diluted antibody, to check we were observing the expression of *env* protein rather natural melanin staining.

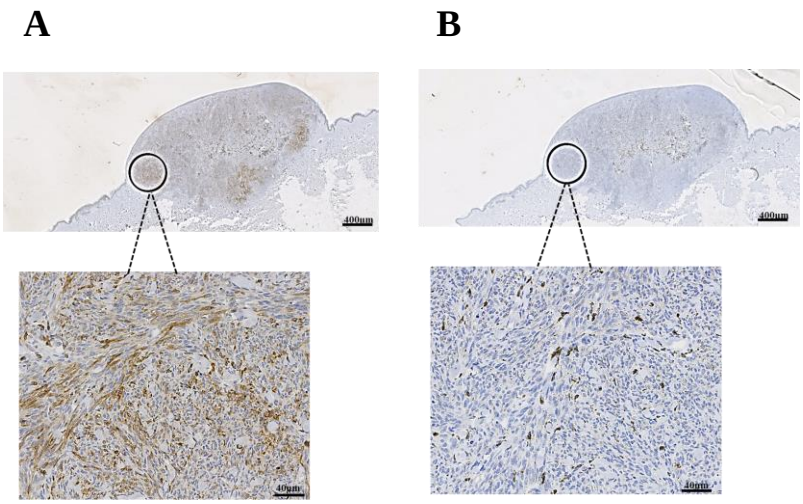


Figure 4.15: Expression of HERV-K *env* protein in representative melanoma tissues.

(A). Intense cytoplasmic and moderate nuclear staining of *env* protein was detected in metastatic melanoma tissue. **(B).** negative control of *env* antibody in the same section of tissue. Images were captured by using Hamamatsu Slide Scanner microscopy (Magnification 10x and 40x)

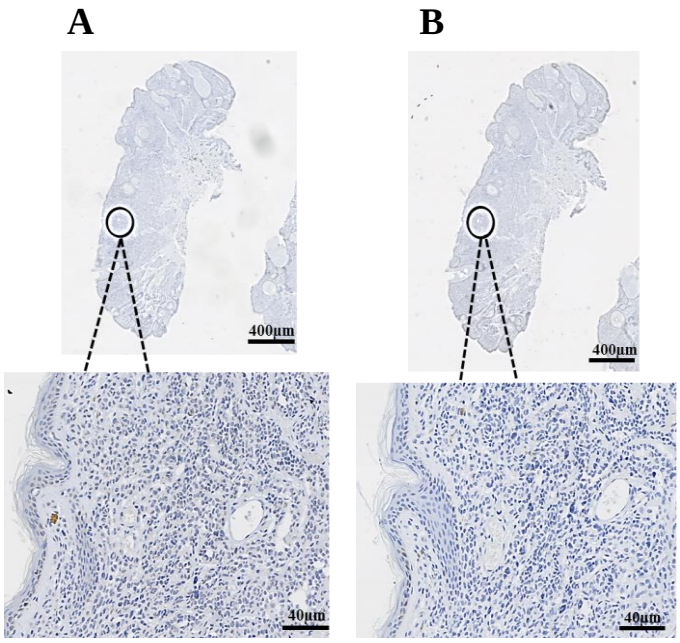


Figure 4.16: Expression of HERV-K env protein in nevus tissues.

Staining of *env* protein was not detected in nevus tissue compared to melanoma. B negative control of *env* antibody in the same section of tissue. Images were captured by using Hamamatsu Slide Scanner microscopy (Magnification 10x and 40x).

4.2.6 p-ERK signal is related to expression of HERV K *gag* protein in melanoma cells

BRAF mutation is associated with melanocyte proliferation in melanoma (Bu & Diehl, 2016). At the same time, the Ras/Raf/MEK/ERK pathway is an important regulator of cell signalling in melanoma. ERK is hyper activated in up to 90% of human melanomas by mutational activation of growth-factor receptors (Willmore-Payne *et al*., 2005).

It has been noted that inhibition of ERK phosphorylation, via treatment with oral doses of specific inhibitors, in patients with metastatic melanoma, showed an 81% response for treatment metastatic melanoma patients, indicating that B-RAF kinase activity plays an important role in BRAF-mutant melanoma proliferation (Bollag *et al*., 2010).

Prior studies have noted the importance of HERV-K activation in the growth of melanoma cells, and an altered BRAF-MEK-ERK pathway is an early event often observed in melanoma progression (Li *et al.*, 2010). These findings, led us to explore whether HERV-K expression is directly regulated by the BRAF-MEK-ERK pathway in the cell lines studied.

The correlation between HERV-K *gag* protein expression, and phospho-ERK (p-ERK) detection, was tested using the p-ERK inhibitor (UO126) at two concentrations: 10 um/ml and 20 um/ml. A panel of melanoma cells were cultured in 10% FBS, and the p-ERK inhibitor was applied to the cells after 24 hours with one well of the six well dish treated with DMSO lacking UO126, as a control. Strikingly the phenotype of the cells changed when treated by the p-ERK inhibitor, as illustrated in Figure (4.17).

In examining the morphology of melanoma cells, we observed more protrusions and extension of the cellular membranes compared to vehicle treated (DMSO) control cells. To examine the expression of HERV-K *gag* protein, cell lysates were collected at different p-ERK inhibitor concentrations, as well from control wells and analysed by western blot. Our data showed that expression levels of *gag* were diminished by the p-ERK inhibitor especially at the higher concentration of 20 um/ml. The quantification of protein level revealed that A375P, A375M, SKMel5 and SKMel28 showed significantly suppressed HERV-K *gag* protein expression when treated with p-ERK inhibitor (Figure 4.18). It is clear from our data that HERV-K *gag* expression is associated with activation of the p-ERK signalling pathway in melanoma cells, as its inhibition reduces *gag* protein levels. These findings may have clinical relevance, because the correlation of signalling pathways (p-ERK) and HERV-K *gag* expression may open a new perspective to unravel the etiology of melanoma.

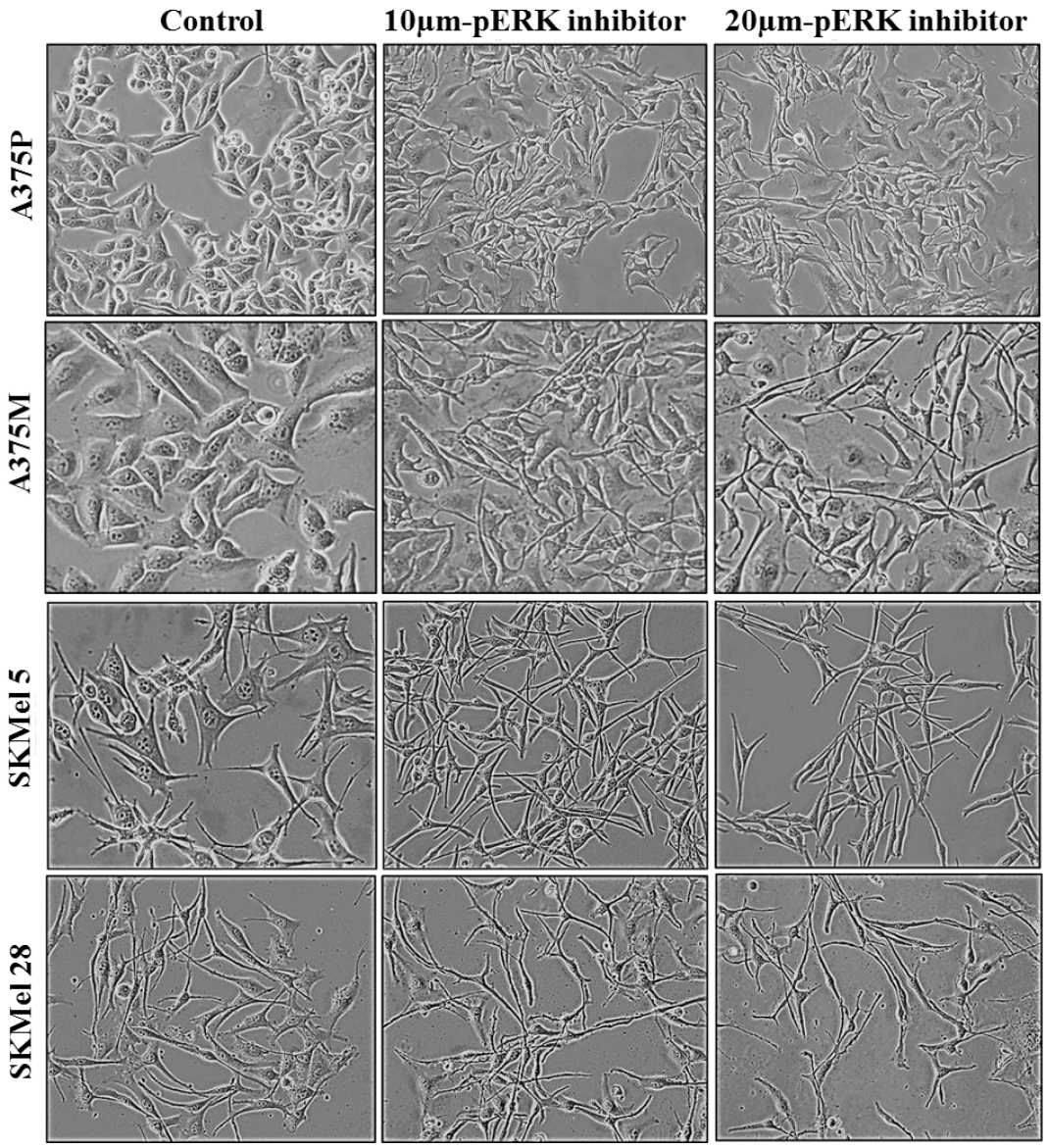


Figure 4.17: Morphology of melanoma cells cultures with p-ERK inhibitor.

Phase contrast images were taken with a 20x objective, after 72 hours. Morphological changes were evident, and especially pronounced with 20 µm/ml of pERK inhibitor.

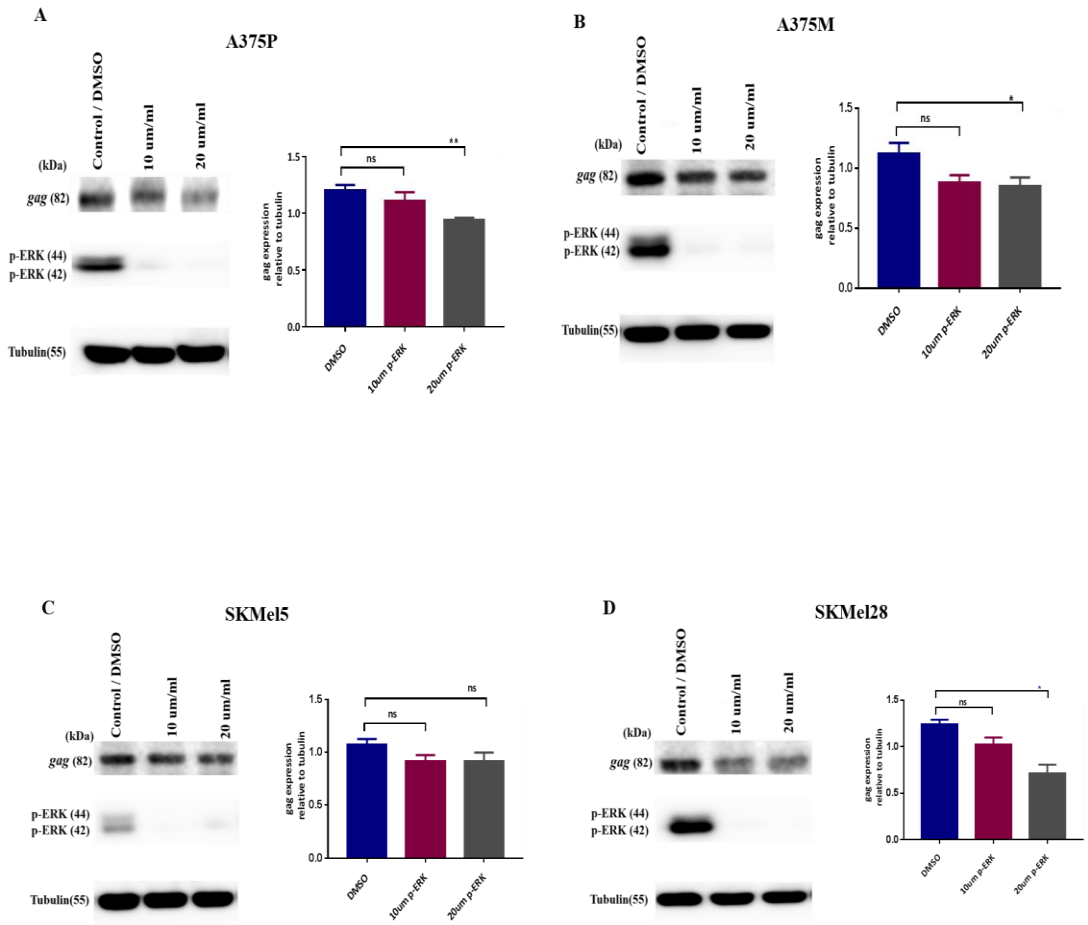


Figure 4.18: Detection HERV K *gag* protein in a panel of melanoma cell lines treated with p-ERK inhibitor.

Melanoma cells were grown in 10% FBS. Cells were lysed and proteins analysed via western blotting. Blots were probed with a primary antibody against *gag*. p-ERK inhibitor (UO126) was added at 10 and 20 um/ml. α- Tubulin was used as protein loading control. The expression of *gag* protein was suppressed in all melanoma cell lines, especially with 20 um/ml. Representative images of multiple replicate experiments (n = 3) are shown. The bar shows quantification level of melanoma *gag* protein, using ImageJ software. Our data showed the *gag* protein was significantly down regulated in all melanoma cells. **(A)** A375P, **(B)** A375M, **(C)** SKMel5 and **(D)** SKMel28.The values plotted are means ± SEMs by one-way ANOVA. Non-significant (ns),* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$ vs Control.

4.3 Discussion

In the first part of this work, HERV-K *gag* and *env* proteins levels were detected in melanoma and other cancer cells in two culture conditions by using western blotting, immunofluorescence and immunohistochemistry assays. Unfortunately, equivalent experiments for the *pol* protein could not be carried out, as no *pol* monoclonal antibody is commercially available.

4.3.1 Investigation of *gag* and *env* proteins in melanoma cells and other cancer cells under two conditions

Our results indicated that *gag* protein expression was evident in all melanoma cell lines prostate cancer cell lines (LNCAP), colon cell lines (SW480), except for the breast cancer cell line T47D, when cultured in media containing 10% FBS. However, under 1% FBS all the cancer cell lines showed significant downregulation of *gag* expression with the exception of LNCAP which did not show any significant change, which is not consistent with activation of HERV provirus gene expression under serum starvation stress, which has been observed by other researchers.

With regards to HERV-K *env* protein level, all melanoma and other cancer cell lines exhibited different expression levels. With T47D showing amongst the strongest expression under 10%FBS. Contrary to expectation, the application of serum starvation (1% FBS) resulted in most melanoma cell lines and two unrelated cancer types (A375P, A375M, SKMel28, LNCAP and SW480) down regulating *env*. (Figures 4.5 and 4.6). The most surprising aspect of the data is that the SKMel5 cell line showed an increase in *env* proteins expression and under 1% FBS culture, but specifically in the adherent, rather, than suspension cells. This suggests that transient expression of *env* may

correlate with the change in phenotype observed in some melanoma cells lines, but is not required for its maintenance.

Basically, there was a significant difference between the two culture conditions but different cell lines showed different responses. This result may be explained by the fact that starvation conditions may change some interactions between cellular signalling pathway such as the MEK-ERK pathway, which are connected to *gag* and *env* protein expression, as observed by Li *et al*. (2010).

Another possible explanation for these changes in gene expression is that the DNA methylation of HERV-K LTR may be increased in 1% FBS in some cancer cells, consequently reducing expression of *gag* and *env* proteins. A study by Lavie *et al*. (2005) found that CpG methylation caused promoter silencing at the HERV-K LTR, but it remains to be seen whether the rapid changes in proviral expression could be the result of DNA based epigenetic modification, given the short timescales.

With regards to SKMel5, surprisingly this cell line showed increased expression of *env* protein under 1% FBS, at least amongst adherent cells compared, with cells in 10% FBS. However, the *gag* protein was gradually decreased in expression (Figure 4.7). It is possible, therefore, that the stress condition may increase expression of HERV-K in one melanoma cell line, but, it is not a general phenomenon for all melanoma cell lines, questioning its relevance to clinical situations.

Contrastingly, Serafino *et al*. (2009) demonstrated a positive correlation between changes in the phenotype of their patient derived clonal melanoma cell line and increased expression of HERV-K *env* proteins. In addition, inhibition of *env* expression

was shown to revert the change in phenotype. Our results are consistent in terms of the potential for *env* expression to respond to serum starvation, but the lack of detection of *env* expression in suspension phenotype cells, argues this protein is not maintaining the phenotype, at least in the cell line SKMel5.

Our analyses also showed that melanocytes had significantly lower expression of the *gag* protein compared with melanoma cell lines (Figure 4.10). More importantly normal primary melanocytes did not express *env* protein, when compared with melanoma cells (Figure4.12). This is consistent with the *env* overexpression seen in clinical studies of a majority of malignant melanoma and provides a tractable experimental system for further study of the *env* protein as a biomarker or even as a therapeutic target in the treatment of melanoma.

This result is in agreement with the *in vivo* data obtained by Buscher *et al*. (2005), where they found *gag* and *env* proteins were expressed in melanoma cells, while normal neonatal melanocytes did not express *env* protein. Another study demonstrated that the melanoma cell lines (SKMel28 and SKMel1) and melanoma tissues expressed *env, rec* and *np9* proteins (Buscher *et al*., 2006).

Generally, our data showed that prostate, colon and breast cancer cell lines showed different patterns of HERV-K protein expression under normal culture conditions (Figure 4.6), so these results will be discussed in more detail.

Our results showed that the LNCAP prostate cancer cell line did express the *gag* and *env* proteins, but only weakly, particularly when compared with melanoma cell lines. This is consistent with studies where HERV-K *gag* or solo LTR expression was

reported in prostate cancer cell lines (including LNCAP) as well as and patients with prostate cancer (Agoni *et al*., 2013; Reis *et al*., 2013; Goering *et al*., 2015). Also, (Rastogi *et al.*, 2016) found a relatively higher expression level of *gag* protein in patients who have prostate cancer, compared to healthy individuals. ERG (ETS-related gene) protein was upregulated 50% of prostate cancer patients, it is possible that these proteins may be associated with the development of prostate cancer. This study also found that ERG protein expression was a strongly correlated with expression of HERV-K *gag* protein in patient who has prostate cancer. These findings suggest that the LNCAP cell line could be a useful model to dissect the role of HERV-K genes expression in prostate cancer and that the *gag* protein could be a biomarker for diagnosis of prostate cancer.

There is little published data on colon cancer and HERV-K gene expression, although one study mentioned that the transcription factor hepatocyte nuclear fator1 (HNF-1) bound to a HERV LTR which acts as promoter for β1, 3-glactosyltransferase5 (β3Gal-T5), a gene that is active in colon cancer. It was argued that this situation led to the development of colon cancer in patients (Dunn *et al*., 2003). Another study reported that HERV-R *env* protein was expressed in colon cancer tissue, compared with normal tissue but they found the env protein expression did not associate with the degree of colon cancer (Lee *et al.*, 2014).

With regards to T47D cells, our data showed that the HERV-K *env* protein was highly expressed in 10% FBS, and significantly increased in expression under 1% FBS. By contrast, this cell line did not express the *gag* protein in either condition. This data suggests that the gag and env proteins, despite sharing a promoter are decoupled with respect to expression level, which might reflect altered splicing or protein stability.

Also a recent study by Lavie *et al.* (2005) found the unmethylated promoter of HERV-K was silent in T47D cells, apparently because of a lack of appropriate transcription factors. The raises the question of what promoter drives *env* expression in T47D. Recent studies have reported a positive correlation between HERV-K *env* expression and breast cancer cells in normal culture conditions (10% FBS) and patients who have breast cancer. HERV-K *env* mRNA was significantly increased in blood from breast cancer patient relative to healthy individuals (Wang-Johanning *et al.*, 2014). Moreover, the transcription of *env* was significantly decreased in breast cancer patient who were treated with chemotherapy, in contrast to primary stage of breast cancer patients, again suggesting potential as a potential biomarker for breast tumours (Rhyu *et al.*, 2014). Furthermore, the knockdown of *env* expression in breast cancer cells was observed to significantly decrease invasion and migration in mouse models and decreased cancer growth and size (Zhou *et al.*, 2016).

Our western blot results were confirmed by immunofluorescence, showing the expression of *gag* and *env* proteins was localized to both the cytoplasm and nucleus (Figure 4.9). In contrast, (Muster *et al.*, 2003) reported *env* protein expression was in the cytoplasm only. With regards to IHC results, HERV-K *env* protein expression was detected in two out of 4 melanoma tissue compared to nevus tissue (0/4), which is consistent with previously reported rates of *env* expression in clinical studies(Zhou *et al.*, 2016).

Buscher *et al.* (2006) detected HERV-K transmembrane envelope (*env* TM) protein in 8 of 21 melanoma tissue samples and the *rec* protein was detected in 13 of 21 samples. While they did not detect np9 protein in the same tissue. They generated polyclonal antibody by using goat and rabbit sera. In contrast, Muster *et al.* (2003) observed

HERV-K proteins in all primary melanoma (9/9) and metastates (3/3) However, HERV-K *gag*, *env* and *rec* proteins were expressed in only one sample at the nevus stage (1/25).

Our data provide insights into *env* protein expression for many reasons. Firstly, normal primary melanocytes did not express the *env* protein, compared with melanoma cells. Secondly, 1% FBS serum starvation can increase expression of this protein in SKMel5 cells that retain an adherent phenotype. Finally, *env* protein expression is significantly increased in T47D cells under 1% FBS compared with 10% FBS. Thus, it could be used as biomarker for melanoma and breast cancer and may be a target for therapeutic approaches.

4.3.2 p-ERK signal is related with expression of the HERV K *gag* protein in melanoma cells

The second part of this work investigated how the pERK signalling pathway was related with HERV-K *gag* expression, and changes in the morphology of melanoma cells (Figure 4.15 and 4.16).

Our results showed that the phenotype of melanoma cells was modified by treatment with a p-ERK inhibitor – the cells appeared to show longer extensions of the cell body compared with control, especially at a concentration of 20 um/ml. This could be related with changes to the signalling pathways that are important for maintaining their cellular morphology. One concern is whether, the toxicity of the pERK inhibitor used may be associated with a change in phenotype. However, a recent study by Cavener *et al.* (2010) found the inhibitor pERK we used showed limited toxicity in pancreatic cancer cells, partially addressing these concerns. Our data showed that HERV-K *gag*

expression is likely activated by the pERK signalling pathway, because the expression of *gag* was significantly decreased in melanoma cells, exposed to 20 μm/ml of pERK inhibitor.

A recent study by Riverso *et al.* (2017) reported the pERK inhibitor caused reduced levels of the mRNA and protein of Kruppel-like factor 4 (KLF4), which is highly expressed in melanoma cells (A375, SKMel5 and SKMel28). Interestingly, Li *et al.* (2010) demonstrated that HERV-K *gag* and *env* protein levels were associated with activation of the pERK signalling pathway in melanoma cells, more than in neval cells by IHC. In addition, they found HERV-K *env* protein was decreased when using p-ERK inhibitor in A375 cells, HERV-K *env* protein was even more reduced by a combination with inhibitors of P16/NK4A/CDK4, by western blotting.

Other authors studied leukaemia cells, where HERV*np9* was observed to be important for regulating ERK1/2 signalling pathway. For example, when the authors silenced *np9* expression, pERK, c-myc and Notch1 pathway activity were significantly decreased levels in leukaemia cells (Chen *et al.*, 2013). In our results, it was clear that *gag* protein expression was regulated by p-ERK. Taken together, all the results above suggest, that HERV-K proteins may have great clinical relevance in cancer and could help to design novel therapeutic target to control them. However, further studies are needed to understand the molecular mechanisms in detail.

The next chapter, therefore, moves on to discuss HERV-K methylation levels within melanoma cells and other cancer cell lines in different culture conditions, and whether this epigenetic modification at the HERV-K promoter correlates with observed changes in mRNA and protein levels.

5 Studying epigenetic modification of the HERV K 5′ LTR in melanoma cell lines

5.1 Introduction

Within human genomic DNA (gDNA) 80% of CpG dinucleotides regions are methylated at the 5′ cytosine residue, with the exception of clusters of CpGs (CpG islands) that are located near or in the promoter of ubiquitously expressed genes (Benesova *et al.*, 2017). This methylated status of CpG rich regions has impact on the expression of genes by repressing transcription, by reducing access of the RNA polymerase to the DNA, presumably by chromatin structure alteration (Menendez *et al.*, 2004).

Particularly in cancer, it has been noted that the change in global methylation levels in gDNA contributes to changes in gene expression. For example, hypomethylation of CpG regions has been found in cells from ovarian, prostate and melanoma cancers (Benesova *et al.*, 2017). It has also been noted that hypomethylation of mobile elements is associated with many cancers (Florl *et al.*, 1999; Lavie *et al.*, 2005).

With regards to human specific mobile elements, L1 elements, Alu elements and HERVs have all been reported to show low levels of DNA methylation in ageing, embryogenic and cancer cells, and this hypomethylation generally correlates with elevated expression of these elements (Sukapan *et al.*, 2014). In contrast, in the early stages of colon cancer, some tumor suppressor genes acquire a high level of methylation (hypermethylated status), and repression of expression of these may be associated with the development of colon cancer, for example (Baylin *et al.*, 2001).

The mechanism and consequences of hypermethylation and hypomethylation on HERV-K expression, are illustrated in Figure 5.1.

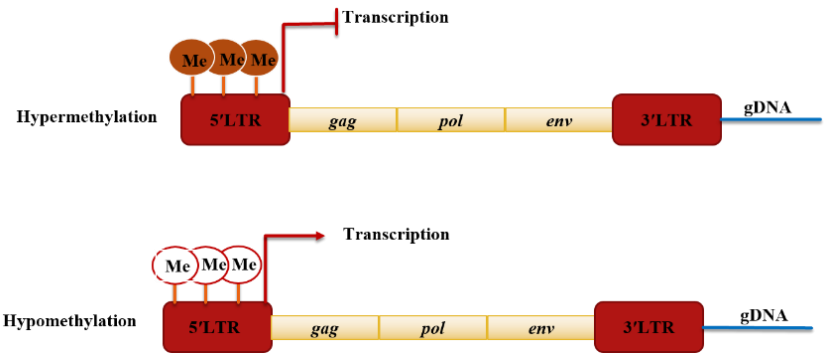


Figure 5. 1: Diagram of methylation status in HERV 5′ LTR promoter.

Hypomethylation / hypomethylation of the HERV 5′ LTR has been shown to directly regulate the transcriptional activity of HERV-K.

Groudine *et al.* (1981) showed that ERV genes can be activated and expressed after treatment by 5-Aza-2′-deoxycytidine (5AzadC), a DNA methylation inhibitor. Moreover, Stengel *et al.* (2010) found that melanoma cells did not express HERV-K transcripts of mRNA such as *env* and *rec* genes due to methylation of the HERV-K 5′ LTR promoter, as treatment of these cells with 5AzadC, triggered HERV-K expression.

This chapter, I will discuss the correlation between modification of epigenetic status, specifically DNA methylation at CpG dinucleotides, of the HERV-K 5′ LTR region in gDNA of melanoma cell lines and HERV-K expression mRNA or proteins. In outline, DNA methylation status was determined in melanoma cells and melanocytes under two culture conditions (10% FBS, 1% FBS), where cells showed different adherent and suspension growth phenotypes. The next step of the investigation was to use, the 5AzadC DNA methylation inhibitor to alter global hypomethylation in melanoma cell lines. 5AzadC acts by being incorporated into gDNA and inhibited DNA methyltransferases (DNMTs) enzymes that bind to it (Hurst & Magiorkinis, 2017). A modification of a Combined Bisulphite PCR and Restriction Assay (COBRA) analysis (Stengel *et al.*, 2010) was used to assess DNA methylation in the HERV-K 5′ LTR; the

restriction step was replaced with direct sequencing. This assay was applied to melanoma cells grown at two different concentrations of 5AzadC (2 uM and 10 uM).

5.2 Results

5.2.1 Amplification global methylation genome DNA of HERV- K 5′ LTR in melanoma cells using COBRA

To determine the contribution of DNA methylation status in controlling HERV-K expression in melanoma cells, DNA methylation levels in the 5′ LTR in four melanoma cells (A375P, A375M, SKMel5 and SKMel28) and normal primary melanocytes were measured. The COBRA assay, reported previously for analysis of HERV-K DNA methylation was applied, but sequencing of cloned PCR products replaced the restriction digestion, to enable analysis of all the CpGs within the amplified region, from HERV-K provriruses across the genome. The principle of the COBRA assay is that sodium bisulphite treatment of gDNA converts ummethylated cytosines to uracil while maintaining methylated cytosines. PCR amplification leads to replacement of uracil with thymine, enabling the inference of the original status of the nucleotide by DNA sequencing.

In this experiment, bisulphite conversion of genomic DNA from cells grown in 10% FBS and 1% FBS was used to determine the methylation levels of CpG dinucleotides contained within 5′ LTR. The modified DNA was subjected to PCR amplification with HERV-K 5′ LTR specific primers. A 3 µl aliquot of PCR product was run on 3% agarose gels and the expected amplicon (of 222 bp) was obtained in all melanoma cells, which contains up to 10 CpG dinucleotides, as illustrated in Figure 5.2. The remaining PCR products were ligated and transformed into *E. coli* and plasmid clones isolated. 24 clones were selected for each cell line from 10% FBS and 1% FBS growth conditions for

sequencing. Each sequenced clone revealed the level of non-CpG cytosine conversion to uracil, to assess the efficiency of bisulphite conversion, as well as the presence of methylated CpGs.

These processes are illustrated in Figure 5.2. Statistical analysis of methylation variation in cloned bisulphite PCR products of all cells was done using the QUantification tool for Methylation Analysis (QUMA) software (Kumaki *et al.*, 2008). Only samples showing a bisulphite conversion efficiency of > 96% were used in the analyses.

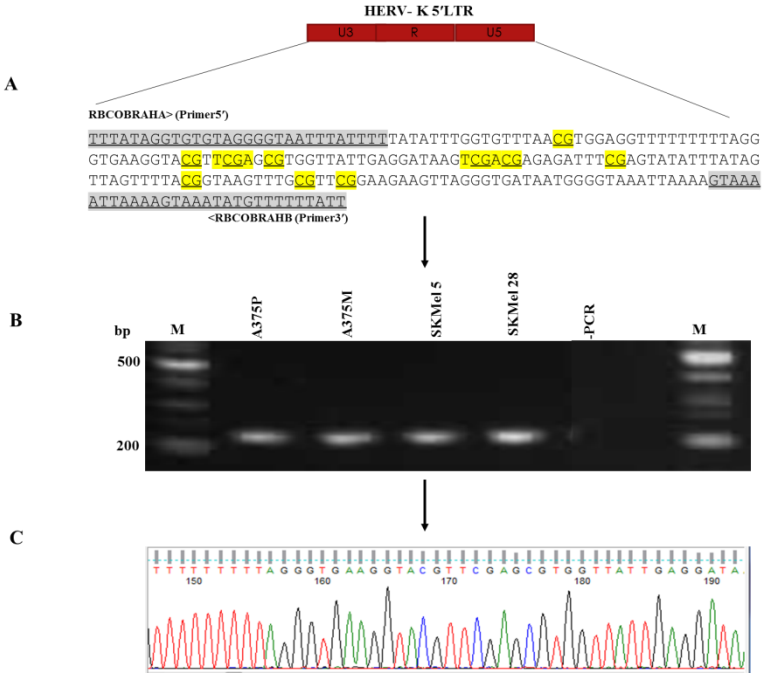


Figure 5. 2: Amplification of bisulphite treated methylated DNA of HERV-K 5′ LTR.

(A) The sequence of a fully CpG methylated genomic DNA derived from the HERV-K consensus sequence map (all non-CpG Cytosines converted to thymine), illustrating the specific primers to that amplify the HERV-K 5′LTR, across the genome. **(B)** Electrophoresis of PCR products from the 5′ LTR (222bp), amplified from bisulphite-modified DNA from melanoma cells under 10% FBS growth conditions. NEB 100 bp marker was used. **(C)** An example sequence trace from a cloned 5′ LTR amplicon. The black arrow indicates a methylated cytosine. Conversion of non-CpG cytosines is complete – only cytosines in a CpG (methylated) context are observed.

5.2.2 Analysis methylation levels of HERV- K 5' LTR promoter in melanoma cells

We aimed to investigate the global DNA methylation levels in HERV-K 5′ LTR promoters in melanoma cells and melanocytes, versus the reference sequence (Figure 5.2 A). Up to 24 clones of bisulphite PCR products from gDNA were analysed by the QUMA software, for each cell line, under 10% FBS. Our data showed that all melanoma cell lines were hypomethylated compared with the reference 5′ LTR sequence, with few cloned amplicons showing complete methylation (Figure 5.3, black circles). It is likely that hypomethylation of the 5′ LTR could lead to the HERV-K promoter becoming active and initiating transcription. However, by contrast in melanocytes, the 5′ LTRs of HERV-K elements were relatively hypermethylated.

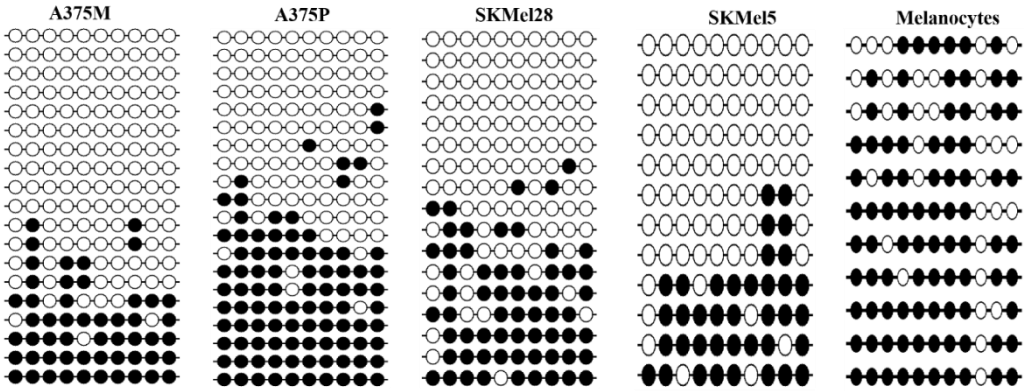


Figure 5. 3: The level of DNA methylation of HERV-K 5′LTRs in melanoma cell lines and melanocytes.

Bisulphite sequencing analysis of 5′ LTR regions revealed all melanoma cells under 10% FBS condition were hypomethylated. Circles represent 10 methylated (black) and unmethylated (white) CpGs within HERV-K5′ LTR. The results of sequence analyses of 11-19 independently derived cloned molecules are presented. DNA hypermethylation of 5′ LTR regions was observed in melanocytes.

5.2.3 Statistical analysis of 5′ LTR promoter methylation level in melanoma and melanocytes

To determine whether methylation levels in melanoma and melanocytes cells were statistically different or not, up to twenty-four cloned sequences were analysed by Mann-Whitney U-test, within the QUMA software. However, the small number of sequenced clones represent only a small sample of HERV-K LTRs in the whole genome and therefore provide a limited assessment of CpG methylation in these elements. Our data showed that the difference of methylation between the melanoma and melanocytes cells was significant for all but A375P (Figure 5.4). The most interesting aspect of these charts is the consistent hypomethylation of the HERV-K promoter compared to melanocytes. The p values of the Mann-Whitney U-test, across all 10 CpG site analysed were A375P (0.05), A375M (0.001), SKMel5 (0.002) and SKMel28 (0.002). Altogether, the above results suggest that the methylation level of the HERV-K 5′ LTR promoter is consistent with the increased level of HERV transcription often observed in melanoma cell lines.

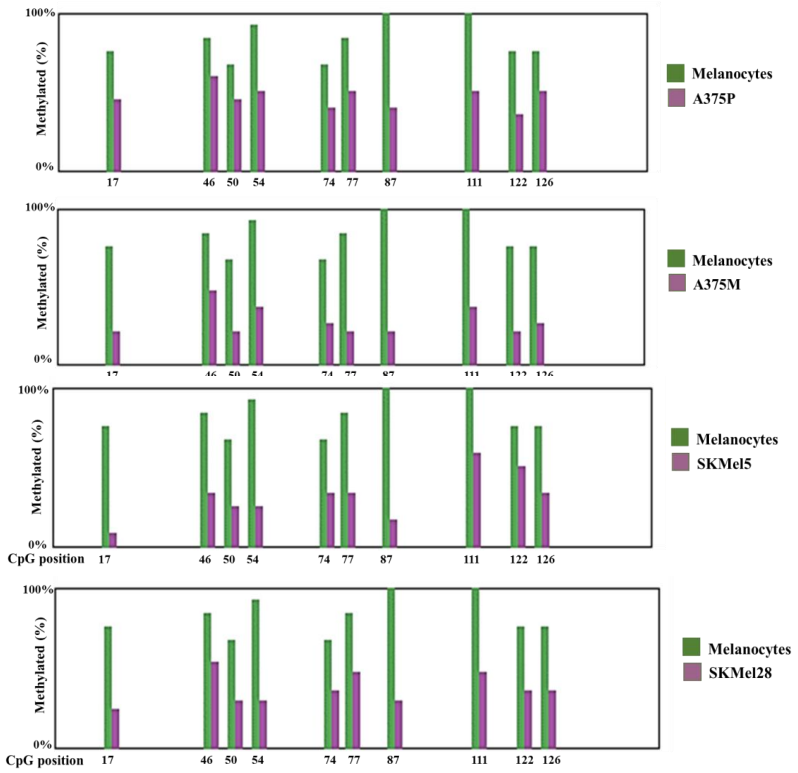


Figure 5. 4: Comparison of CpG methylation between melanoma cells and melanocytes.

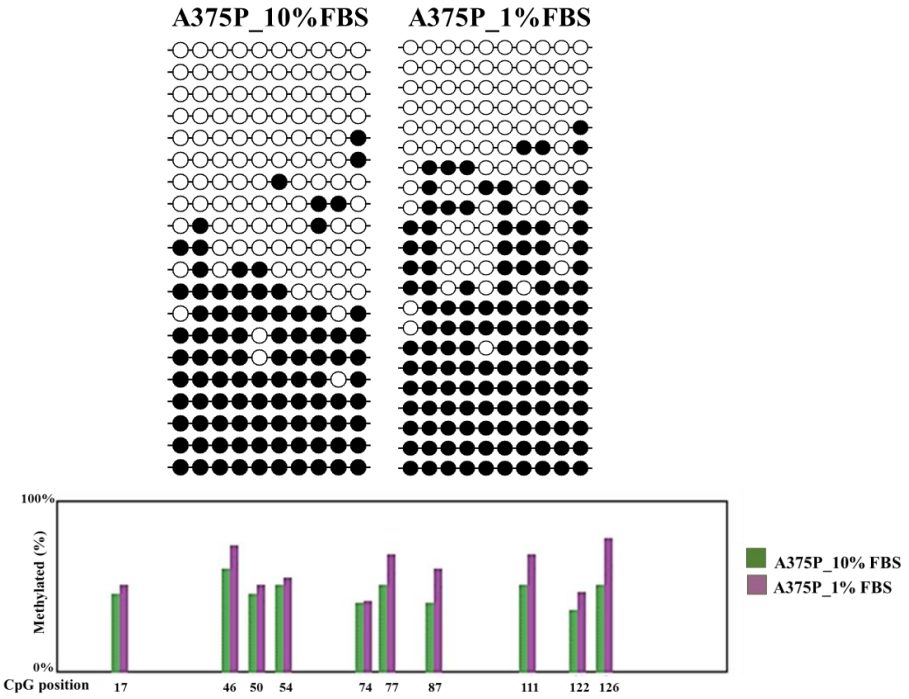
The bars chart show the methylation levels of 5′ LTR promoter in A375P, A375M, SKMel5 and SKMel28 compared with melanocytes, using QUMA software on cloned bisulphite PCR product sequencing data.

5.2.4 Comparison of CpG methylation 5′ LTR in melanoma cells between two conditions

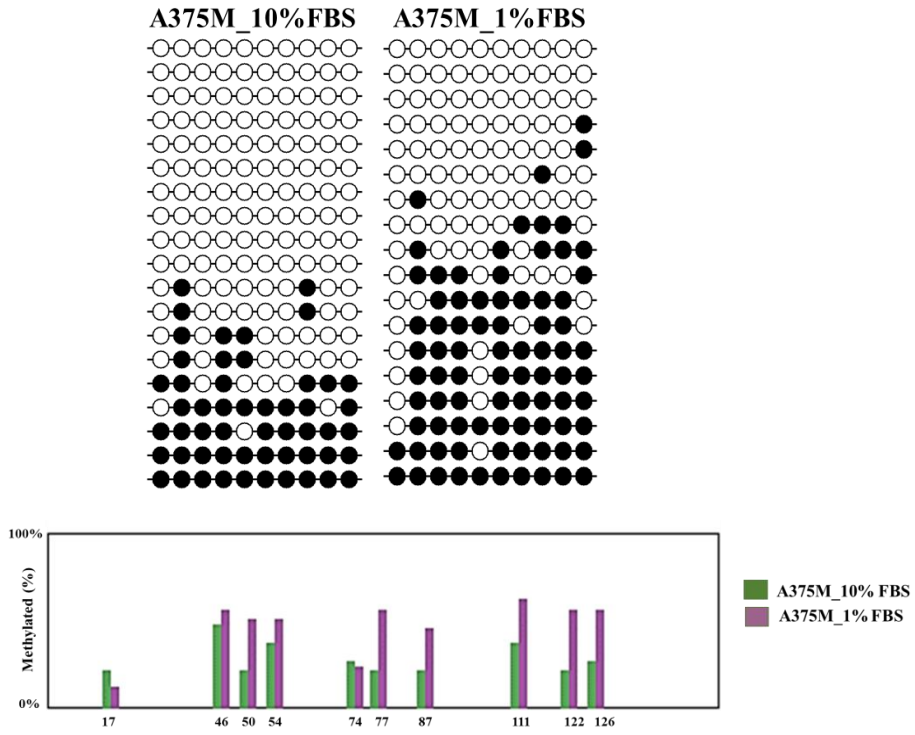
We then asked whether methylation levels changed in melanoma cells under normal 10% FBS and 1% FBS (serum starvation) culture conditions. To address this question, gDNA was extracted from melanoma cell lines grown under the two different conditions and cloned bisulphite PCR amplicons were analysed. Our data showed that melanoma cells which were grown under 10% FBS recorded a low level of methylation, between (20-40%). By comparison the same, melanoma cell lines (A375P, A375M, SKMel28) growing as adherent cells in 1% FBS, showed a consistently increased level of CpG methylation, (Figure 5.5 A-C). However, on comparison with the Mann-

Whitney U-test, no significant difference between the two conditions was evident. The p values of 10% versus 1% growth conditions were A375P (0.3), A375M (0.1), SKMel28 (0.08).

A



B



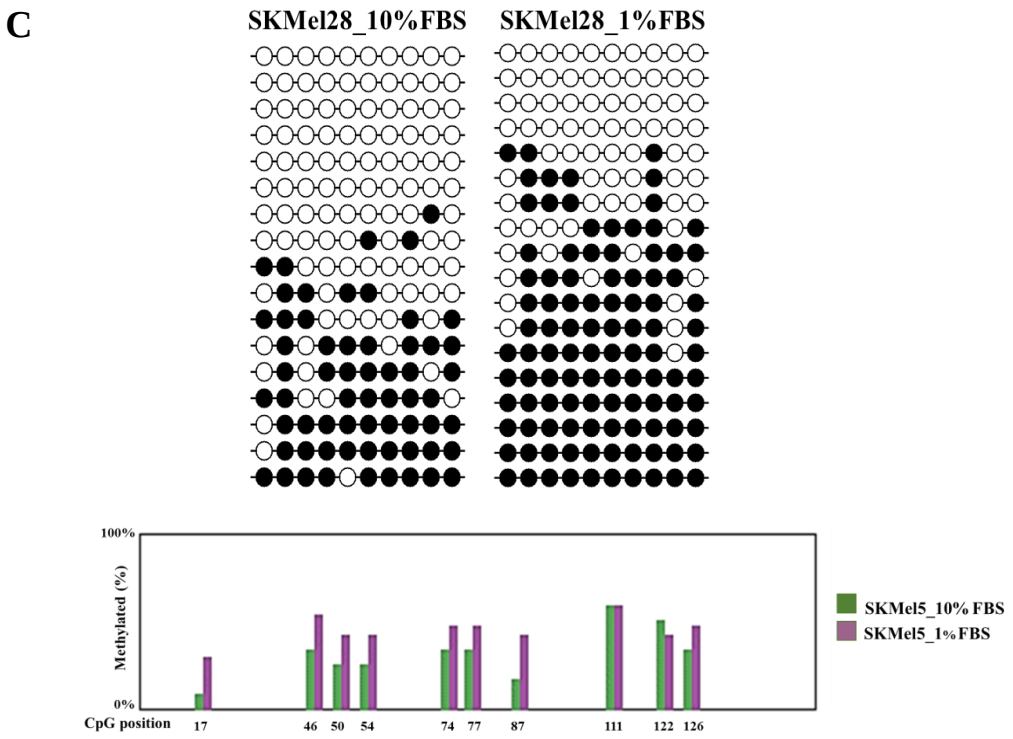


Figure 5. 5: Comparison of CpG methylation of melanoma cells between two conditions.

Bisulphite sequence diagrams and bar charts showing the methylation levels of 5′ LTR promoter CpGs in **(A)** A375P, **(B)** A375M, and **(C)** SKMel28 which were grown in 10%FBS compared to the same cells in 1% FBS. The QUMA software was used to analyse cloned bisulphite PCR product sequencing data.

Next, we sought to investigate the SKMel5 cell in 1% FBS, growing as adherent and suspension cells, to see whether any changes in methylation level correlated with the cells' phenotype. The bisulphate sequence analysis showed an increase in methylation level under 1% FBS condition for both cellular phenotypes (adherent and suspension) compared with SKMel5 grown in 10% FBS (Figure 5.6 A). Moreover, the methylation level was increased more under 1% FBS in suspension cells relative to cells growing as adherent in 10% FBS (Figure5.6 B). However, despite this consistent increase in the level of DNA methylation, no significant difference between the two conditions was evident, by Mann-Whitney U test: the p value of comparing SKMel5 cells in 10% FBS and SKMel5 cells in 1% FBS, growing in adherent morphology was (0.09), while

SKMel5 in 10% FBS versus SKMel5 1% FBS in suspension morphology, it was (0.2). Our results suggest that there is some influence of serum starvation on DNA methylation levels, and that the trend is for DNA methylation to increase under serum starvation. A possible explanation for this might be increased activation of DNA methyltransferase enzymes under 1%FBS, although the mechanisms behind this are not clear.

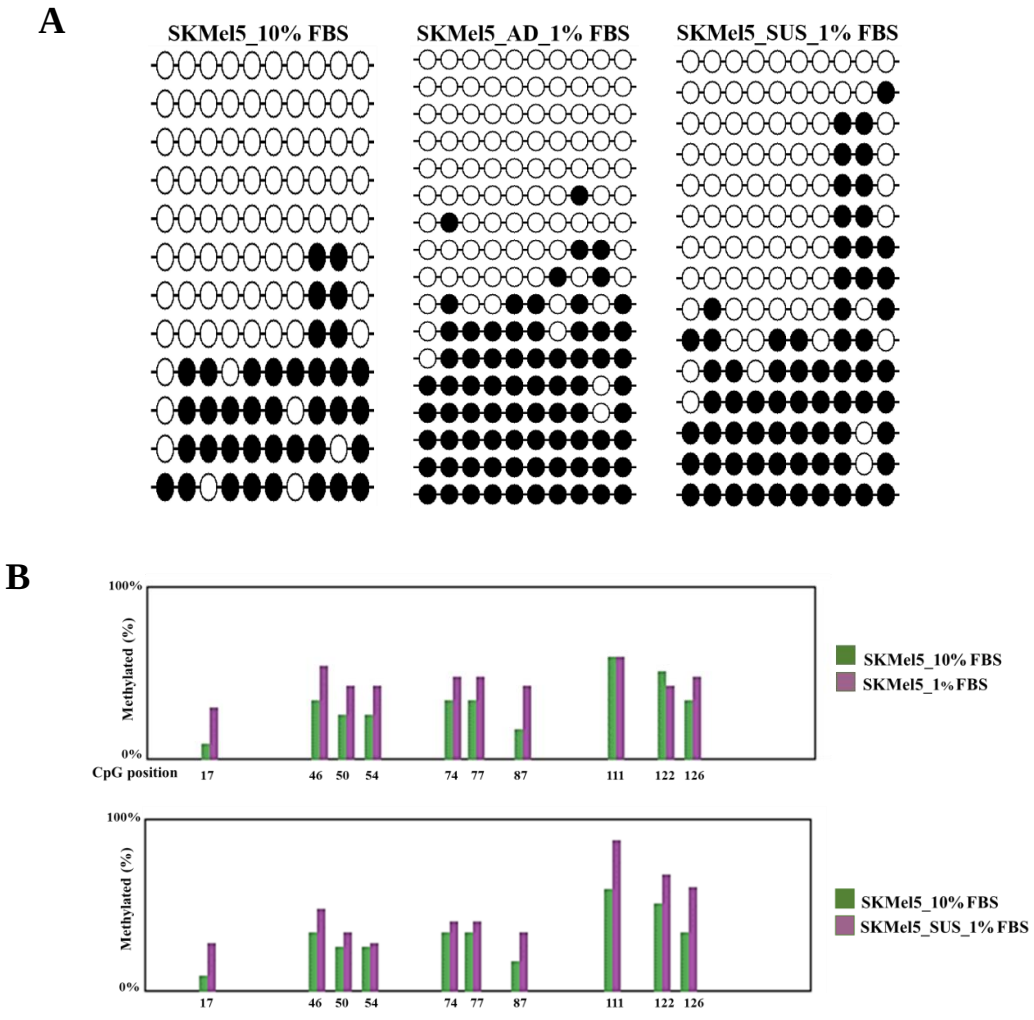


Figure 5. 6: Analysis of methylated DNA of HERV-K 5′LTR in the SKMel5 cell line.

(A) Bisulphite sequencing analysis of the 5′ LTR regions in SKMel5 revealed that methylation level under 10% FBS condition were low. Under 1% FBS DNA methylation increased according the phenotyping, **(B)** The bar charts show the methylation levels of the 5′ LTR region in SKMel5 which was grown in 10%FBS compared with 1% FBS condition (as adherent cells) and suspension cells (SUS). The QUMA (Quantification Unite Methylation Analysis) software was used to analyse cloned bisulphite PCR product sequencing data.

5.1.1. The effect of 5-Aza-2′-deoxycytidine (5AzadC) on expression of HERV-K proteins in melanoma cells

As previous experiments showed that while melanoma cell lines are significantly hypomethylated at HERV-K 5′ LTRs, relative to normal melanocytes, serum starvation, which has been reported to activate HERV expression, in our hands produces consistent increases in global DNA methylation. To investigate further whether DNA methylation influences HERV expression, we analysed DNA methylation and HERV gene expression in cells exposed to the DNA methylation inhibitor, 5AzadC. To assess the effect of the inhibitor melanoma cell lines A375M, SKMel28, and normal melanocytes as a control, were grown on 10% FBS. After 24 hours, these cells were treated with 5AzadC, at a concentration of 10 µM, for 72 hours. An inhibitor negative control (using only DMSO, the solvent in which 5AzadC is dissolved) was carried out for each cell line. Then, DNA extraction, bisulphite PCR and cloning and sequencing was used to measure DNA methylation at the HERV-K 5′ LTR. Fisher's Exact test, as implemented in the QUMA software, was used to compare the level of DNA methylation between treated and control cells: SKMel28 and A375M in 10 µM 5AzadC showed significant reductions in DNA methylation ($p=0.008$ and 0.02 respectively) compared with control (DMSO only) (Figure 5.7). However, melanocytes did not show significant differences in DNA methylation when treated with 10 µm of 5AzadC for 72 hours (Fishers Exact test, $p=0.5$).

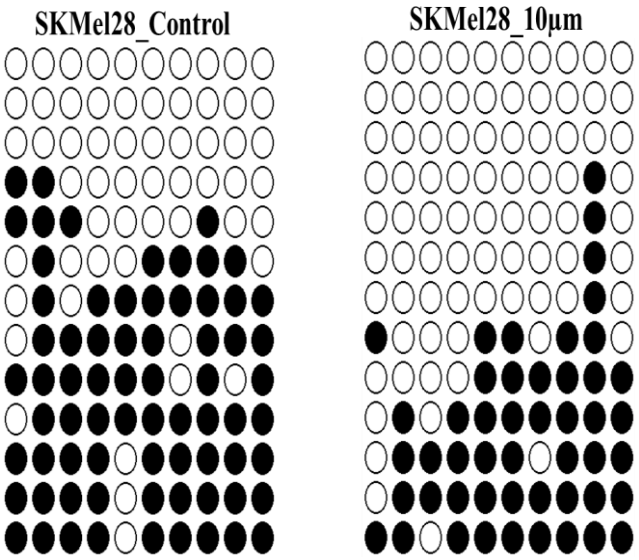


Figure 5. 7: Example of 5AzdC inhibitor reducing DNA methylation.

Bisulphite sequence analysis of HERV-K 5′ LTR regions revealed SKMel28 cells treated with 10 µm of inhibitor were hypomethylated relative to control. The overall percentage of unmethylated CpGs following by 5AzadC) in SKMel28 was 62.6% was higher than in the untreated SKMel28 control 55.3%. Black and white circles represent methylated and unmethylated CpGs respectively.

Next, 5AzadC treatment was used with three melanoma cell lines A375P, A375M and SKMel28 grown in two serum conditions (10% FBS and 1% FBS). After 24 hours, these cells were treated with 5AzadC, at 2 different concentrations (2 µM and 10 µM) for 72 hours, followed by western blotting. This experiment included controls where cells were treated with just DMSO, under the two growth conditions.

We were interested in whether 5AzadC treatment, although evidently effective in reducing HERV-K 5′ LTR DNA methylation, might alter HERV protein expression as a result of de-repression of the HERV promoter. To address this, we carried out western blotting with specific *gag* and *env* monoclonal antibodies. Cell lysates from three melanoma cell lines, which were treated and grown in different conditions, were used. 20 µg of proteins were loaded in all western blot experiments. Western blot analysis showed that *gag* protein expression was decreased compared with control in both

growth conditions (10% FBS and 1% FBS) (Figure 5.8 and Figure 5.9). However, it is evident that at the higher concentrations of inhibitor (10 μM), *gag* expression was increased in A375P and SKMel28 under 10% FBS, suggesting that at least for these cell lines promoter hypomethylation may increase HERV expression in melanoma cells and other cancer cell lines, but not to higher than control levels.

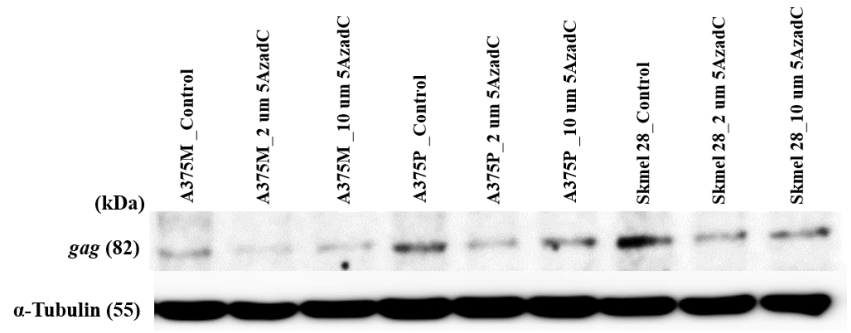


Figure 5. 8: Detection of HERV K *gag* protein in 10% FBS condition with 5AzadC treatment.

Melanoma cells were grown in 10% FBS. Cells were treated by 5AzadC at two different concentrations. The cells were incubated for 72 hours and lysates of these cells analysed via western blotting. Blots were probed with antibody against *gag*. α- Tubulin was used as protein loading control. The expression of *gag* protein was down regulated in all melanoma cells. Representative image of multiple replicate experiments (n = 3).

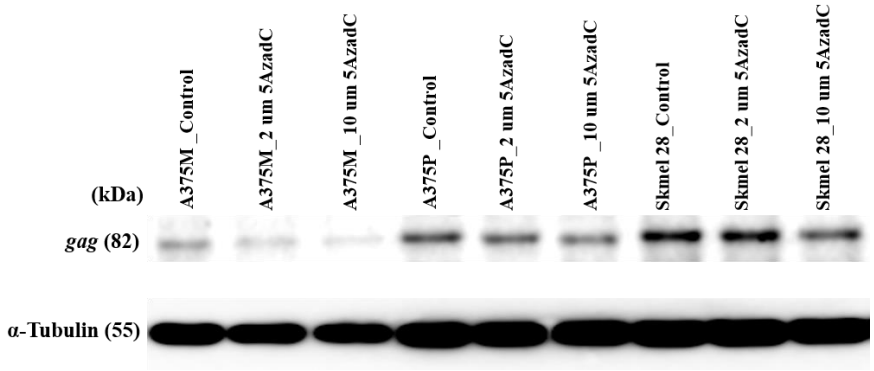


Figure 5. 9: Detection HERV K *gag* protein in 1% FBS condition with 5AzadC treatment.

Melanoma cells were grown in 1% FBS. Cells were treated by 5AzadC at two different concentrations. The cells were incubated for 72 hours and lysates of these cells analysed via western blotting. Blots were probed with antibody against *gag*. α- Tubulin was used as protein loading control. The expression of *gag* protein was down regulated in all melanoma cells. Representative images of multiple replicate experiments (n = 3).

Quantification of HERV-K protein levels, using *gag*, was analysed under the two growth conditions and two inhibitor conditions. The analysis showed that melanoma cells significantly decreased expression of *gag* protein after treatment with 5AzadC in 10% FBS at both inhibitor concentrations. However, *gag* expression in 10 µM 5AzadC was significantly increased compared to 2 µM 5AzadC (Figure 5.10). In 1% FBS culture all cell lines showed significantly decreased expression of *gag* protein at 10 µM of 5AzadC, while there was no significant difference in A375P and SKMel28 with 2µM treatment under 1% FBS culture (Figure 5.11).

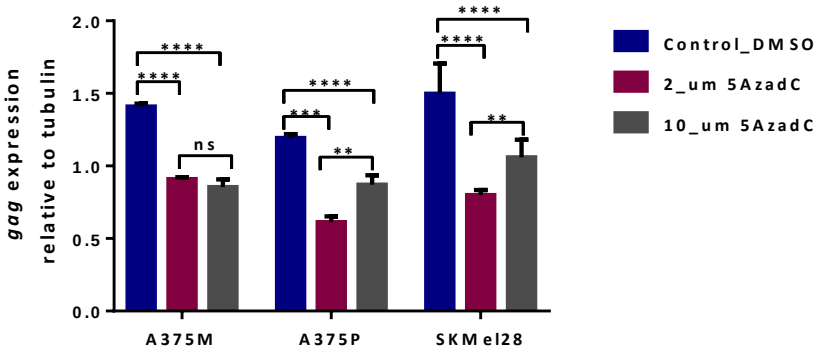


Figure 5. 10: Quantification of *gag* protein in 10% FBS conditions.

The bar chart shows the *gag* protein levels quantified using ImageJ software according to the western blot results. The results showed the 10µM of 5AzadC impacted *gag* expression. The values plotted are means ± SEMs by two-way ANOVA followed by Student's unpaired t test between two groups. Non-significant (ns), * p< 0.05, ** p< 0.01, *** p< 0.001, **** p< 0.0001 vs Control.

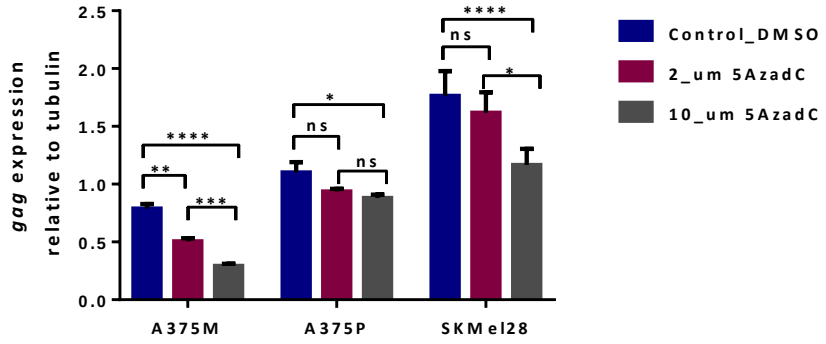


Figure 5. 11: Quantification level of *gag* proteins in 1% FBS condition.

The bar chart shows the *gag* protein levels quantified using ImageJ software according to the Western blot results. The results showed the 10µm of 5AzadC impacted *gag* expression. The values plotted are means ± SEMs by two-way ANOVA followed by Student's unpaired t test between two groups. Non-significant (ns),* p< 0.05, ** p< 0.01, *** p< 0.001, **** p< 0.0001 vs Control.

Next, we sought to investigate the HERV K *env* protein expression in the same melanoma cells in 10% FBS and1% FBS. Lysates from cells, grown under the two levels of FBS and treatments with 5AzadC. The same assay was carried out to observe any changes of *env* protein after 72 hours of inhibitor treatment. Western blot analysis revealed that A375P, A375M and SKMel28 decreased expression of the *env* protein when compared with control, under two conditions (Figure 5.12 and Figure 5.13).

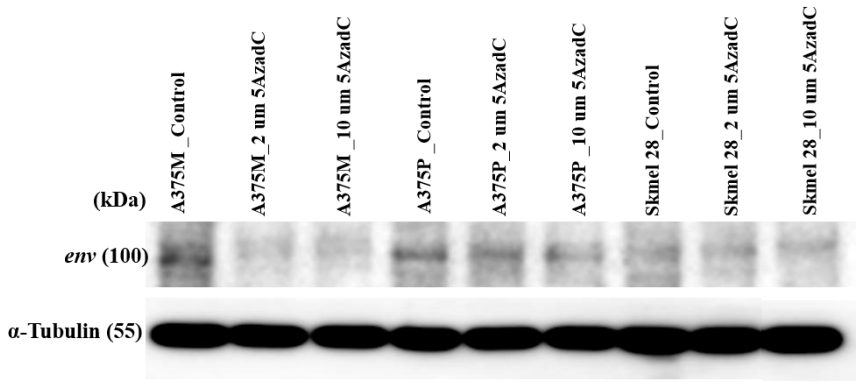


Figure 5. 12: Detection HERV K *env* protein in 10% FBS condition with 5AzadC treatment.

Melanoma cells were grown with 10% FBS. Cells were treated by 5AzadC in two different concentrations. The cells were incubated for 72 hours and lysates of these cells collected. Western blots were probed with an antibody against *env*. α- Tubulin was used as protein loading control. The expression of *env* protein was down regulated in all melanoma cells. Representative image of replicate experiments (n = 3).

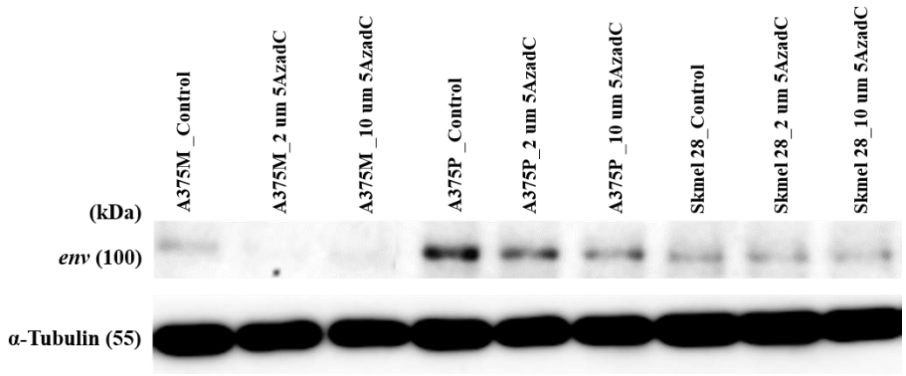


Figure 5. 13: Detection of HERV K *env* protein in 1% FBS conditions with 5AzadC treatment.

Melanoma cells were grown with 1% FBS. Cells were treated by 5AzadC in two different concentrations. The cells were incubated for 72 hours and lysates of these cells collected Western blots were probed with antibody against *env*. α- Tubulin was used as protein loading control. The expression of *env* protein was down regulation in all melanoma cells. Representative images of replicate experiments (n = 3).

Quantification of the western blot data showed that A375P and M cell lines significantly decreased expression of the *env* protein in 10% FBS and1% FBS cultures, compared with control. Interestingly there was no significant difference in SKMel28 under the two

growth conditions (Figure 5.14 and Figure5.15). This suggests that, for this cell line *env* protein expression was not affected by 5AzadC treatment, despite the fact that 10 uM 5AzadC significantly reduces DNA methylation at the HERV-L 5′ LTR in SKMel28 cells (Figure 5.14).

In conclusion, the present results demonstrated that HERV K *env* and *gag* proteins in melanoma cells under 10% FBS and 1% FBS conditions showed decreases in presence of. 5AzadC treatment. However, the level of methylation was not related with expression of HERV-K proteins after 5AzadC treatment, despite its demonstrable effect on promoter methylation, indicating that there is not a simple, direct relationship between promoter methylation and protein expression.

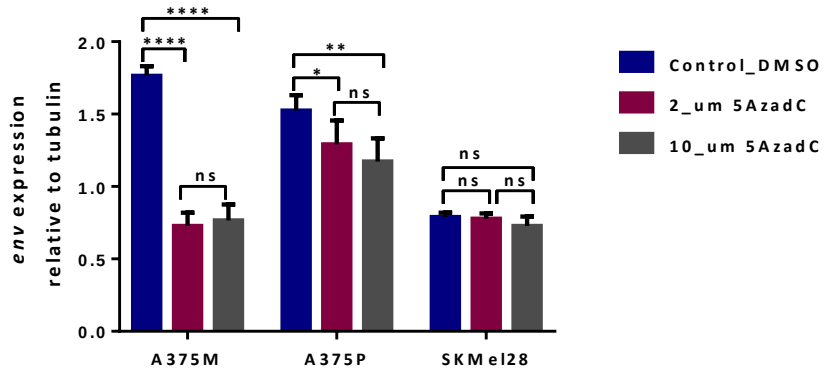


Figure 5. 14: Quantification level of env proteins in 10% FBS condition.

The bar chart shows the *env* protein levels quantified using ImageJ software according to the Western blot results. The results showed the 10µm of 5AzadC was impact on *env* expression. The values plotted are means ± SEMs by two-way ANOVA followed by Student's unpaired t st between two groups. Non-significant (ns),* p< 0.05, ** p< 0.01, *** p< 0.001, **** p< 0.0001 vs Control.

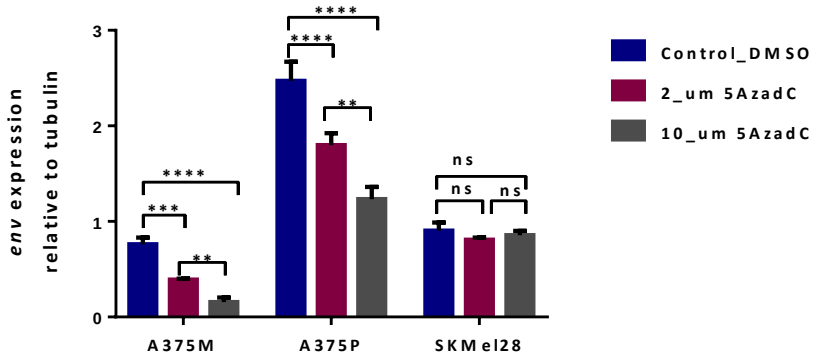


Figure 5. 15: Quantification of env proteins in 1% FBS condition.

The bar chart shows the *env* protein levels quantified using ImageJ software according to the western blot results. The results showed 10µm of 5AzadC impacted *env* expression. The values plotted are means ± SEMs by two-way ANOVA followed by Student's unpaired t test between two groups. Non-significant (ns), * p< 0.05, ** p< 0.01, *** p< 0.001, **** p< 0.0001 vs Control.

5.2.5 The effect of 5-Aza-2′-deoxycytidine (5AzadC) on expression of HERV-K proteins in melanocytes

While melanoma cell lines show DNA hypomethylation, congruent with the tumours they are derived from, our results indicate that this epigenetic modification is decoupled from HERV protein expression. However, melanoma cell lines are adapted to culture and are highly sensitive to DNA methylation inhibitors. We next investigated whether 5AzadC inhibitor affected HERV expression in normal primary melanocytes (HEMaLP cells). After 24 hours of culture, these cells were treated with 5AzadC, testing 2 different concentrations of 5AzadC (2 µM and 10 µM) for 72 hours, followed by western blotting. This experiment was included negative controls (DMSO only).

Gag and *env* protein expression was determined by western blot. Protein quantification showed that *gag* protein was only decreased significantly in 10 uM 5AzadC, but it was a small effect (Figure 2.16 A and B). Surprisingly the expression of *env* was not detected in melanocytes before and after treatment with 5AzadC (Figure 2.16C).

It can therefore be assumed that the 5AzadC inhibitor did not dramatically change the activity of the 5′ LTR HERV-K, and that the results observed for melanoma cell lines did not hold for melanocytes.

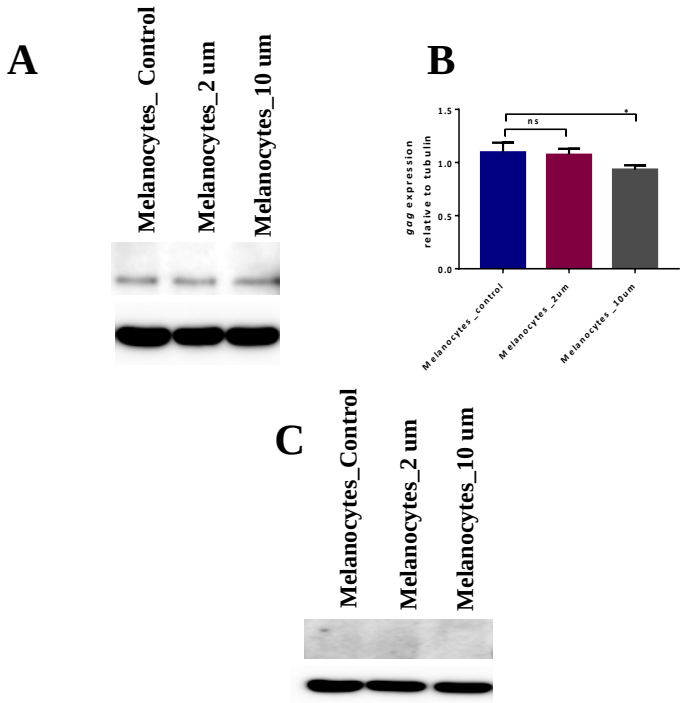


Figure 5. 16: Detection HERV K *gag* protein in melanocytes cell with 5AzadC treatment.

(A) Melanocytes cell was grown with 10% FBS. Cell was treated by 5AzadC in two different concentrations. The cell was incubated for 72 hours and collected lysed of these cells. Protein analysed via western blotting, blots were probed with antibody against *gag*. α- Tubulin was used as protein loading control. The expression of *gag* protein was down regulated in melanocytes. Representative images of replicate experiments (n = 3). **(B)** The bar chart shows the *gag* protein levels quantified using ImageJ software according to the Western blot results. The results showed that only10µm of 5AzadC impacted *gag* expression. The values plotted are means ± SEMs by two-way ANOVA. Non-significant (ns),* p< 0.05, ** p< 0.01, *** p< 0.001, **** p< 0.0001 vs Control. **(C)** *Env* protein analysed via western blotting, blots were probed with antibody against *env*. α- Tubulin was used as protein loading control. The expression of *env* protein was not detected in melanocytes. Representative images of replicate experiments

(n = 3).

5.3 Discussion

In the present chapter, methylation of HERV-K 5′ LTR promoter was investigated in a panel of melanoma cells (A375P, A375M, SKMel28 and SKMel5), as well as normal melanocytes. To determine the relationship between methylation levels in 5′ LTR promoter region and HERV-K protein expression. However, the number of sequenced clones was a small sample, and may not be representative of methylation levels of HERV-K LTR in gDNA of melanoma cells. This process involved establishing a modified COBRA assay to quantify DNA methylation in a CpG rich region of the consensus HERV-K 5′ LTR, assessment of the influence of culture conditions and DNA methylation inhibitor treatment on DNA methylation, and analysis of the consequences on HERV-K protein expression by western blotting.

5.3.1 Determination of methylation levels in melanoma cells

As mentioned above, in previous chapters, melanoma cell lines were shown to express mRNA from the HERV-K *pol* gene, the *gag* protein and also weakly expressed the *env* protein. Melanocytes were used as a control and were shown to express *pol* mRNA, and the *gag* protein, although at a lower level than melanoma cell lines. Finally, we could not detect the *env* protein in melanocytes.

As the *gag* protein (along with *pol* and *pro*) and the *env* protein are encoded by alternatively spliced transcripts, initiated from the same promoter, these assays enabled us to address the factors influencing HERV-K genes expression in terms of epigenetic regulation and protein production. This foundation led us to investigate how manipulation of culture conditions (serum starvation) and the inhibition of DNA methylation (with 5AzadC) impacted HERV expression in our panel of melanoma cell

lines, compared to normal melanocytes. Both of these treatments have been reported to elevate HERV expression in various cell types (Roulois *et al*., 2015).

With regards to HERV-K DNA methylation results, methylation of the 5′ LTR was observed at a low level in all melanoma cell lines (hypomethylation), according to bisulphite sequence analysis, when cultured in 10% FBS. By contrast, melanocytes (also in 10% FBS) showed the expected high level of DNA methylation when compared to melanoma cells i.e. were hypermethylated (Figure 5.3).

These were statistically significant differences in methylation levels between melanocytes and melanoma cells, and they were consistent with the global demethylation that is characteristic of many cancer cell lines (Figure 5.4). The fact that we observed a pattern of DNA methylation (hypomethylation) at the HERV-K 5′ LTR that may result in relaxation of epigenetic repression of the HERV promoter is consistent with the observation of elevated HERV-K expression specifically in melanoma.

However, despite the fact that melanocytes were hypermethylated at the HERV-K 5′ LTR, they still expressed the *pol* mRNA and the *gag* protein at a low level, suggesting that DNA methylation is not the only factor.

It could be that some transcription factors (TF) used by the HERV-K promoter are not influenced by the methylation status of the 5′ LTR (Stengel *et al.*, 2010).

For example, nuclear transcription factor Y (NFY) and Erythroid transcription factor (GATA-1) have been reported to have the ability to bind the ERV-3 5′LTR promoter even when it is hypermethylated in erythroleukemia cells. However, these authors also measured the relative level of expression of the hypermethylated promoter – this was

much less than the unmethylated ERV-3 5′LTR region, suggesting viral promoters can evade repression by DNA methylation (Hu *et al.*, 2017).

Another possible explanation is that HERV-K is the youngest and least numerous family of ERVs and the deletion or mutation in promoter of HERV of other HERV families results in their activation. A recent study by (Hurst & Magiorkinis, 2017) showed, by microarray analysis, that LTR of the evolutionarily youngest HERV were hypermethylated in normal tissues, while older families of HERVs lost methylation because of sequence divergence through mutations. Furthermore, Alu, HERV-K and HERV-E showed quite different levels of methylation in Systemic Lupus Erythematosus (SLE), compared to healthy individual, suggesting epigenetic regulation of these sequences is specific to the repeat family (Sukapan *et al.*, 2014).

Taken together all results above, the methylation status of the HERV-K 5′LTR is not directly related to the expression of HERV-K in melanocytes.

With regards to HERV-K 5′ LTR methylation of melanoma cells, according to recent reports, demethylation of the 5′ LTR promoter may contribute to the increase in HERV-K expression. The activation of 5′ LTR promoter was decreased after CpG methylation in the Human teratocarcinoma cells (Tera1 cell line,) supporting a direct relationship (Lavie *et al.*, 2005).

In addition, hypomethylation of HERV-K has been shown to play an important role in genomic instability in urothelial cancer (Florl *et al.*, 1999).

It has been reported that some TFs contribute to activation of the HERV-K5′ LTR promoter when it is hypomethylated, especially the SP1 and SP3 factors. Although, the mechanism for these factors action is not clear, SP1 and SP3 have been shown to bind

in four G rich stretches in the HERV-K promoter which act as nucleosome free regions to initiate HERV-K LTR transcription (Hanke *et al.*, 2016)

DNMT enzymes are important for establishment and maintenance of methylation in CpG regions in gDNA, and reduced level of these enzymes are associated with increased HERV-K expression (Stengel *et al.*, 2010). On the other hand, hypermethylation of CpG regions can silence HERV-K expression by preventing TFs binding (Saito *et al.*, 2017), and DNMT enzymes can be inhibited by different ways. For example, 5-Aza-2′-deoxycytidine (5AzadC) inhibitor which has been shown to use in clinical trials. However, 5AzadC inhibiter is not specific for set of genes in gDNA for reactivation. As a result, reactivation of genes will occur randomly because of the decrease in methylation level (Yoo & Jones, 2006). However, Hurst & Magiorkinis (2017) noted that HERVs and their virus-like particles (VLPs) were expressed in Teratocarcinomas, despite their promoters being in a highly methylated state.

In conclusion melanoma cells are hypomethylated at the HERV-K 5′ LTR and show expression of HERV-K proteins, while melanocytes are hypermethylated at the HERV-K 5′ LTR but also show detectable HERV protein expression. However, selection bias is another potential concern because the number of samples were a small to present all methylated LTR promotor in whole gDNA in melanoma cells.

5.3.2 Comparison of methylation levels in melanoma cells under 10% FBS and 1% FBS

All melanoma cells were grown as adherent cells in 10% FBS and 1%FBS cultures and their methylation level at the HERV-K 5′ LTR analysed by bisulphite PCR. In addition, because the cells could be physically separated by their growth phenotype we measured the methylation level of SKMel5 growing in suspension, in1%FBS media. Bisulphite sequence analysis showed that the methylation level of the HERV-K 5′ LTR in A375P, A375M, SKMel28 and SKMel5 with adherent morphology, under 1%FBS conditions, had increased compared to the 10%FBS condition (Figure5.5). However, this increase in methylation level was not statistically significant. In our data from Chapter 3 (Figure 3.10), mRNA expression was decreased under 1% FBS culture, in most cell lines except A375P, and at the same time the expression of *gag* and *env* proteins was significantly decreased. This is consistent with epigenetic modification of the HERV-K 5′ LTR.

A possible mechanism for this observation might be that TF binding may be influenced by promoter methylation states. For example, the YY1 transcription factor may not bind in the HERV-K 5′ LTR if it is methylated. This situation leads to a decrease in HERV-K expression (Kim *et al.*, 2003). Also, methylation could be induced by reactivation of DNMT3A / 3B, which act to add CH_3 group to CpG in LTR regions, as a response to serum starvation (Saito *et al.*, 2017). Moreover, histone modification can play important role in transcription of genes, including methylation and acetylation modification. However, histone modification may act to active or silence genes according to region and type of modification, for example, H3K4 is in activator of promotor regions while H3K9 and H3K27 occur in regions that are repressed for transcription (Kondo *et al.*, 2008).

TP53 expression may have contributed to increase methylation levels and decrease HERV-K expression. Wylie *et al.* (2016) found that in ten patients who had colon cancer, mutation of TP53 was correlated with increasing L1 expression compared to eight normal individuals who harboured wild type TP53. Thus, TP53 apparently restricts expression of L1 in cancer cells and it could be possible that activation of the TP53 tumor suppressor gene is associated with suppression of mobile elements expression.

5.3.3 The effect of 5-Aza-2′-deoxycytidine (5AzadC) on expression of HERV-K proteins in melanoma cells and melanocytes

One proposed mechanism of the AzadC DNA methylation inhibitor's anticancer properties is the reactivation of silenced genes, such as tumour suppressor genes. It has been shown that the AzadC inhibitor plays an important role in checkpoint therapy by blocking the inhibitory effect of check point therapy and then enabling reactivation of immune systems such as interferon responses (INF) to reduce develop cancer in humans. In colon cancer cells, low doses of 5AzadC (inhibitor of DNA methylation) induced viral dsRNA which activated the MDAS/MAVS/TRF7 signalling pathway, which in turn induced INFs and interferon stimulated gene (ISG) that defend against cancer by reducing the proliferation of cancer cells. Therefore, this mechanism can be considered as anticancer immune modulated response (Roulois *et al.*, 2015).

`

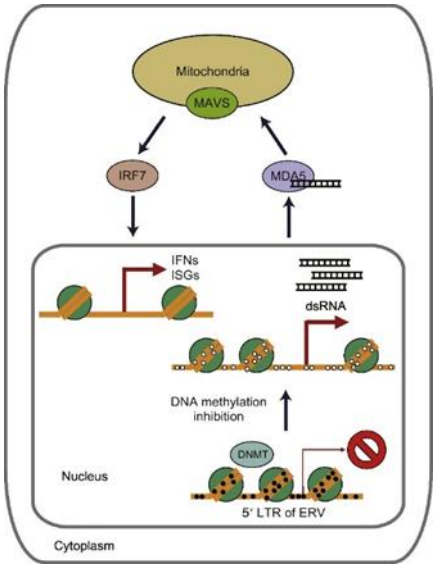


Figure 5. 17: Schematic diagram of IN response by viral ds RNA.

Inhibition of DNA methylation by 5AzadC induced a viral response, mediated by ERVs by inhibition of methylation level in 5′ LTR promotor. MDAS/MAVS/TRF7 signalling induced INF response as antiviral immune protection (Saito *et al.*, 2017)

Our data showed that after treatment with 5AzadC, HERV-K expression was down regulated significantly in melanoma cells and melanocytes. The most interesting finding was that melanoma cells in 10% FBS and 1% FBS cultures showed decreased *gag* (Figure 5.8) and *env* protein expression, (Figure 5.12) especially at the 10 µM concentration.

This result may be explained by the fact that some tumour suppresser genes may act to suppress HERV-K protein expression because of 5AzadC inhibitor mechanism. Wylie *et al.* (2016) reported a negative relationship between mobile elements expression and TP53 expression in three model animals; flies, zebrafish and mice.

Also, Liu *et al.* (2014) mentioned that Histone-lysine N-methyltransferase (SETDB1) silenced ERV expression when the promoter histones were modified. Thus, it could be 5AzadC is working to impact on SETDB1 modification, and consequently decreased expression of HERV-K.

Several reports have shown that small noncoding RNAs (miRNAs) regulate differential and proliferation process in cells by supressing some genes expression, so any change of miRNA levels may lead to the development of cancer (Saito *et al.*, 2017; Calin & Croce, 2006). In a previous study Saito *et al.* (2009) showed that 5AzadC treatment of gastric cancer cell lead to activation of miRNA 512-sp, within an Alu repeat on chromosome 19, resulting in the silencing of the MCL1 gene expression and apoptosis of gastric cancer cells, indicating a direct connection in some situations.

Furthermore, activation of antiviral responses was increased by 5AzadC treatment of cells in organoid culture, apparently by inducing production of dsRNA from murine ERVs (Chiappinelli *et al.*, 2015). Consistent with these findings in mice, Roulois *et al.* (2015) investigated 5AzadC clinical treatment in melanoma patients: melanoma progression was reduced, apparently because 5AzadC induced HERV dsRNA, and interferon signalling was effective in retarding melanoma development. These studies support 5AzadC as an effective DNA methylation inhibitor that can release HERV repression.

In contrast, a previous study demonstrated that melanoma cell lines (SKMel13 and SKMel19 showed no change of HERV-K *env*, *rec* and *np9* genes expression after 5AzadC treatment (Stengel *et al.*, 2010).

With regards to our data, especially on serum starvation and 5AzadC treatment results, it appears that cellular toxicity is not the causative factor of reducing HERV –K proteins

and limitation of cellular resources rather than an epigenetic response. Notably, the western blotting results used tubulin expression as a loading control /normalisation factor, and this protein appeared as high density bands in all experiments that means the cells are clearly alive and making protein.

It was also observed that in 1%FBS condition, the low concentration of 5AzadC resulted in upregulation of both protein levels, compared to 10 uM 5AzadC inhibitor.

In summary, our results showed that HERV-K expression is associated with hypomethylation of the 5' LTR promoter in melanoma cells grown in 10% FBS. However, this is the first study to investigate the effect of serum starvation culture on methylation state in melanoma cells and melanocytes as well as studying HERV-K protein expression. This situation leads us to assume that starvation conditions may impact on some signalling pathways to reduce HERV-K expression. Therefore, the molecular mechanism of reducing protein expression is elusive. Additionally, 5AzadC treatment resulted in down regulation of HERV-K proteins expression.

Taken together, the melanoma cell lines behave the same way when grown in 1% FBS and in 5AzadC inhibitor, resulting in a reduction in HERV-K protein expression. It could be suggested that HERV-K expression was influenced by two mechanisms, methylation of the LTR promoter and activity of transcription factors which are able to initiate transcription even if their binding site in the LTR promotor is methylated.

Further research should be undertaken to investigate the *pol* gene expression by RT-PCR / qPCR / RNAseq to see if the demethylation does indeed increase transcription, despite the loss of protein. It could be that HERV transcription does increase due to promoter demethylation, but some other mechanism prevents translation, such as the interferon response.

6 NGS analysis of HERV-K proviral loci

6.1 Introduction

HERV-K is the youngest family of HERVs, and integrated in human gDNA more than 30 MYA, as intact proviruses (Macfarlane & Badge, 2015). When HERV-K proviruses insert they produce direct repeats of the gDNA, which are called Target Site Duplications (TSDs) which are between 5 and 6 bp long(Kahyo *et al*., 2017). For example, HERVK-113 is flanked by a repeat of 5′-CTCTAT-3′ sequence of DNA on chromosome 19 (Hindorff *et al*., 2009).

Generally, the HERV-K proviral structure is characterised by terminal identical sequences of the LTRs, and between them are located the ORFs, which encode viral proteins. Homologous recombination between LTRs forms solo LTRs with 2500 insertion loci of HERV-K represented as solo LTR. Although most of HERV family are inactive because of deletion, mutation and recombination, HERV-K is still potentially active because nearly 60 HERV-K loci are present as full-length proviruses in human gDNA (Hindorff *et al.*, 2009; Kahyo *et al.*, 2017).

HERV-K expression may be associated with some diseases and its expression is elevated in different types of cancer such as melanoma, breast and prostate (Kahyo *et al*., 2017). However, the mechanism of pathogenesis, or even if their expression is a side effect of disease, is not well understood (Subramanian *et al.*, 2011).

What is clear is that HERV-K insertions are polymorphic in human populations; insertions have occurred over human evolution and can be fixed or lost from the population through drift or selection. According to the hg19 human genome assembly (GRch37/hg19), HERV-K is present at 91 loci as full length proviruses while 944 loci contain solo LTRs (Subramanian *et al.*, 2011). Many methods and techniques are used

to detect polymorphism of HERV-K insertions including bioinformatic analyses, DNA hybridization and suppression PCR (Macfarlane & Badge, 2015; Mourier *et al.*, 2015; Wildschutte *et al.*, 2016).

However, it may be difficult to investigate HERV-K polymorphisms for a number of reasons. Firstly, HERV-K proviruses are comparatively large insertions, nearly 10kb when full length and can be diverged in sequence which can frustrate bioinformatics analysis, particularly those drawing on short read NGS datasets. Secondly, the numbers of samples that can be analysed in a single experiment is limiting, as the amount of HERV sequence is small and whole genome sequencing approaches are thus inefficient. Thirdly, low allele frequency of many insertions, and their population specificity mean exhaustive analysis is impossible (Macfarlane & Badge, 2015; Kahyo *et al.*, 2017). Finally some loci have complex structures (directly repeated insertions are known) or are located in regions of gDNA that are highly repetitive. Thus, some HERV-K loci cannot be readily characterised by current methods.

It has been noted that whole genome next generation sequencing (NGS) is an advanced method to detect HERV insertions as well as to discover novel insertions of HERV-K by producing a huge number of short reads that are mapped to a reference genome. New insertions are indicated by reads containing HERV sequences that map to regions lacking annotated HERVs (Mourier *et al.*, 2015). However, these computational approaches are complex, prone to false positive / negatives and inherently inefficient in terms of bases sequenced.

However, Macfarlane & Badge (2015) demonstrated that provirus specific suppression PCR based, assays Genome-wide amplification of proviral sequences (GAPS) assay, (described in the Material and Methods) designed to initiate amplification from viral *gag* or *env* gene enable selective amplification of proviruses and the identification of novel insertions without requiring whole genome sequencing.

In this paper, they investigated polymorphism of HERV-K proviruses across a panel of human and primate DNA, and discovered a novel insertion in human and chimpanzees. Due to the selective nature of GAPS, and the comparative rarity of HERV insertions, differences in insertion complement can be revealed by suppression PCR and fractionation by agarose gel electrophoresis (Figure 6.1).

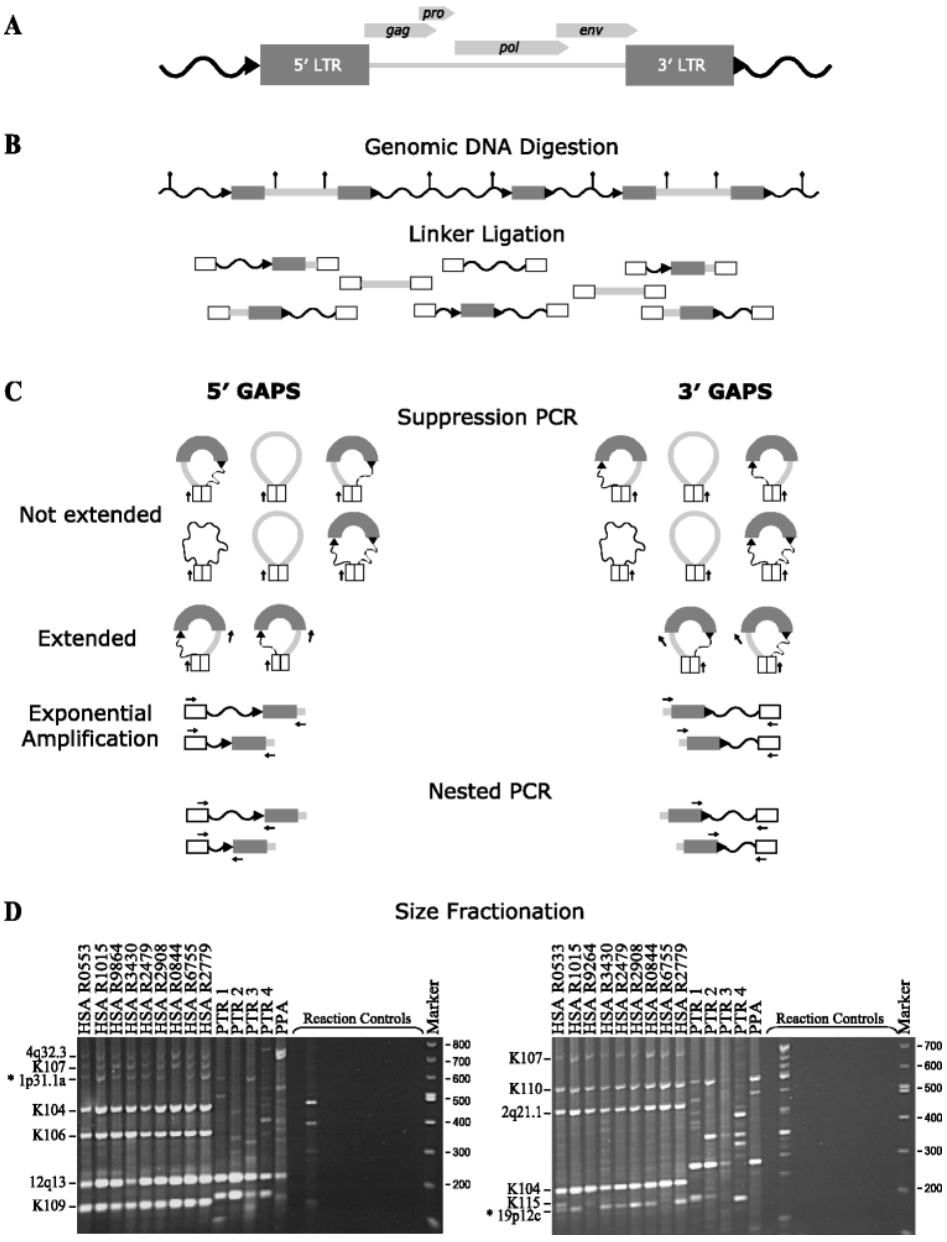


Figure 6. 1: GAPS method for amplification 5′ltr HERV-K and 3′LTR HERV-K from humans and chimpanzees.

(A) Structural of HERV-K included *gag*, *pro*, *pol* and *env* flanked by two LTRs. **(B)** Suppression PCR assay that starts with a linkered genomic DNA library (white boxes). These library fragments serve as templates in suppression PCR. **(C)** GAPS PCR amplification, 5′ GAPS targets the 5′ end, and 3′ GAPS the 3′ end of HERV-K proviruses. Arrows indicate primer binding sites. During suppression PCR, self-complementary linkers form stem-looped DNA molecules that do not amplify. The specific sequence that contains the binding site for the target primer of choice (*gag* or *env*) are extended and become templates for exponential amplification. Nested PCR is performed to enrich for sequences containing a LTR and to reduce amplicon length. **(D)** PCR products are separated by agarose gel electrophoresis - 5′ GAPS on left and 3′ GAPS on right. Genomes include nine humans (HSA) and four chimpanzees (PTR). Lane 14 (PPA) is a positive control. Library controls (left to right): No genomic DNA, no restriction enzyme, no linker ligated genomic DNA, no ligase and no linker. Numbers on the left are DNA size marker fragments (bp). New polymorphic proviruses discovered using GAPS (*). Adapted from Macfarlane & Badge, 2015.

The GAPS assay was designed to avoid amplification of solo LTR and also SVA elements which are short interspersed elements within human gDNA. The (SINE-R) region of SVA shows sequence similarity to the HERV-K 3′LTRs (Deininger *et al*., 2003).

In this chapter, the GAPS assay was used on melanoma cell lines and other cancer cells, including the NTera2D1 cell line which is known to assembly HERV-K VLPs to screen for novel insertions. The GAPS assay was developed to make use of NGS (Ion torrent) so that many samples could be multiplexed in one experiment, and amplicons invisible to agarose gel electrophoresis could be captured. This combination of methods is a novel strategy to detect proviruses with intact *env* genes, by applying the 3′ GAPS assay, and is demonstrably efficient in doing so.

6.2 Results

Genome-wide amplification sequence (GAPS) of HERV-K (3′ LTR) in cancer cells.The GAPS assay is used to selectively amplify a part of HERV-K proviruses from human gDNA and can target the HERV *gag* (5'LTR) OR *env* (3'LTR) regions. This assay avoids amplification of solo LTRs, by requiring presence of the *gag* or *env* genes (Macfarlane & Badge, 2015). Thus, the 3'GAPS assay can be used to characterise the cohort of HERV-K proviruses in the cancer cell lines (A375P, A375M, SKMel5, SKMel28, LNCAP, SW480, T47D and NTera2D) analysed in this thesis. Many of these cell lines express HERV proteins and one, NTera2D1, assembles HERV-K VLPs in cell culture (Bhardwaj *et al.*, 2015). The first step of the analysis was that the 3′GAPS assay was applied to 600ng of gDNA from the 8 cell lines, carried out exactly as described in Macfarlane and Badge (2015). The digested DNA was ligated to the suppression PCR linkers (RBMSL2, annealed to RBD3) and excess linker removed using Millipore centrifugal filter units. Completed ligation of the gDNA was confirmed by 0.8% agarose gel fractionation, as shown in Figure (6.2).

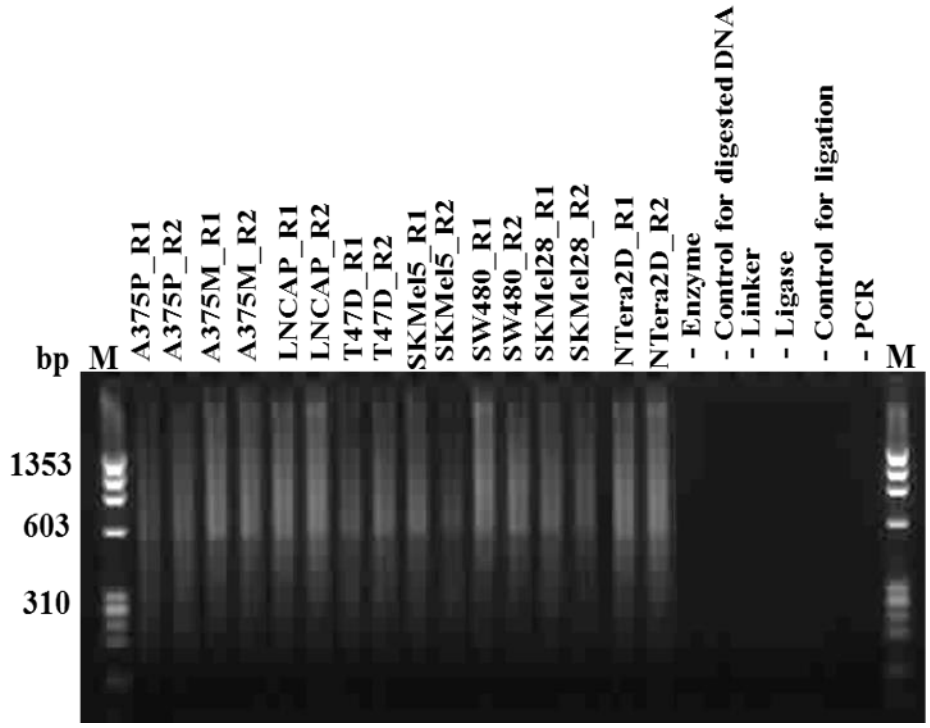


Figure 6. 2: Primary amplification of ligated gDNA of cancer cell lines in 3'GAPS assay.

Two replicates (R1 and R2) of each cell line gDNA were digested with the restriction enzyme *Ase*I and ligated to linkers. In addition to the cell line DNA the experiment included 6 negative controls (-): an enzyme negative control (containing gDNA but lacking the *AseI* enzyme, a negative control (lacking gDNA), a linker negative control (containing gDNA but lacking the linkers), a ligase negative control (containing gDNA but lacking the ligase enzyme), a negative control (lacking digested gDNA for ligated process), and –PCR (where H2O replaced DNA in PCR reaction show any amplification). ϕX174HaeIII marker (M) was used.

The primary suppression PCR was carried out with primers RBX4 (linker-specific primer) and CMKENV (a primer specific for HERVK *env* gene). A secondary nested PCR was set up with primers specific for HERV-K 3'LTRs (CMK3LTR). After thermal cycling, PCR products were fractionated by agarose gel electrophoresis (Figure 6.3). Successful amplification of HERV-K 3'LTRs was indicated by diagnostic differently size bands in each cell lines (Figure 6.3). The gel was southern blotted to transfer the DNA to a nylon membrane and hybridized with a specific probe for the HERV-K107 3'LTR, to confirm specificity of the GAPS assay (data not shown). These experiments

confirmed that 3′ GAPS was able to amplify multiple HERV-K from the genomes of cancer cells.

6.2.1 Variant amplicons

It was noted that the initial fractionation of secondary suppression PCR products showed some cell line specific amplicons, potentially reflecting novel HERV-K provirus insertions. As these might reflect new insertions accumulated during culture of these cell lines or segregating polymorphism they were analysed further. Extra amplicons (different from the canonical human 3′ GAPS pattern) were observed in A375M, SKMel28 and NTera2D. However, we focused on the extra bands from A375M, because this is a derivative of A375P, raising the possibility that this is a novel culture derived insertion, rather than a segregating polymorphism.

In addition, in contrast to previous GAPS experiments, we observed a high molecular weight bright smear in A375P, SW480 and NTera2D secondary suppression PCRs, suggesting alternative amplicons were present. (Figure 6.4). For both the novel discrete amplicons, and the high molecular weight smear, suppression PCR, cloning and sequence were done to investigate.

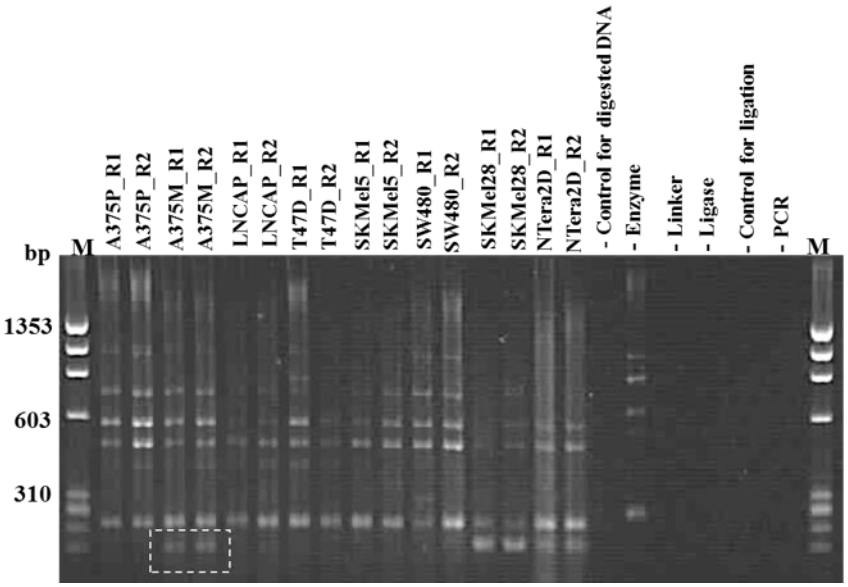


Figure 6. 3: Secondary (nested) PCR Primary.

PCR products were diluted 1:50 in single molecule diluent (5 mM Tris-HCl (pH 7.5), 5 ng/µl sonicated *E.coli* genomic DNA). The diluted amplicons were input to the 3'GAPS assay secondary PCR (35 cycles) using CMK3LTR (primer specific for HERVK 3'LTRs) and RBY1 (primer specific for the linker). The secondary PCR products were size fractionated through 2% agarose gels. An extra band in A375M is shown in the white dotted square. ϕX174HaeIII marker (M) was used. This experiment included negative controls. R refers to replicates.

6.2.2 Investigation of high molecular weight amplicons

These artefacts were investigated as they might reflect failure of the suppression PCR system, or genuine HERV-K proviral amplicons that had previously been ignored. In addition, applying NGS to these amplicon libraries necessarily involves size fractionation and these high molecular weight fractions would be excluded from the sequencing – thus it was important to understand their nature. Cloning and sequencing revealed that the fragments were derived from linkered genomic fragments whose suppression was not effective, and were not enriched for HERV sequences.

As a result, we were confident that their exclusion from the NGS experiment would not result in loss of data. Their occurrence in this GAPS experiment was attributed to the fact the gDNA used was of high quality and high molecular weight, freshly extracted from cells in culture.

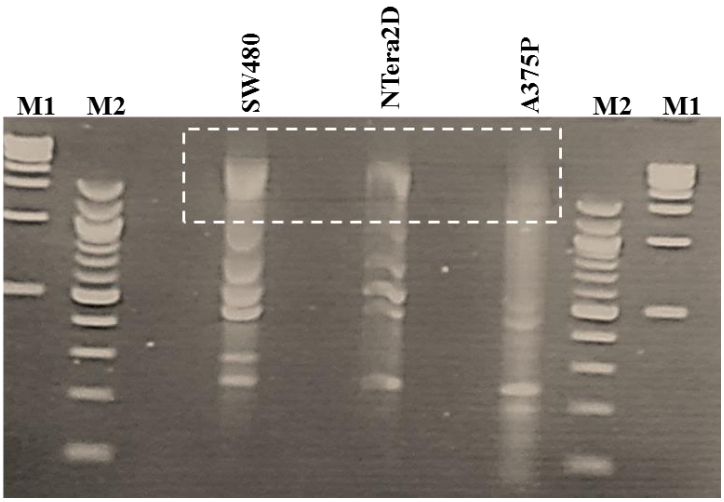


Figure 6. 4: Investigated a bright smear from secondary PCR.

PCR products (from 35 cycles), of SW480, NTera2D1 and A375P were run on agarose gels. After cloning the sequence of amplicons in the bright smear (the white dotted square) was determined. 1Kb (M1) and ϕX174HaeIII marker (M2) were used.

6.2.3 Investigation of extra bands in A375M

The extra band in A375M was cut from the gel, the DNA extracted and cloning and sequencing carried out. The initial results of these sequencing showed the novel amplicon resulted from an *Ase*I restriction site polymorphism in A375M, rather than a novel insertion. To confirm this, we designed a new primer to genotype this polymorphic *Ase1* site. According to the sequence of extra band, the new primer should amplify a product of nearly 350 bp, with the CMK3LTR primer (Figure 6.5)

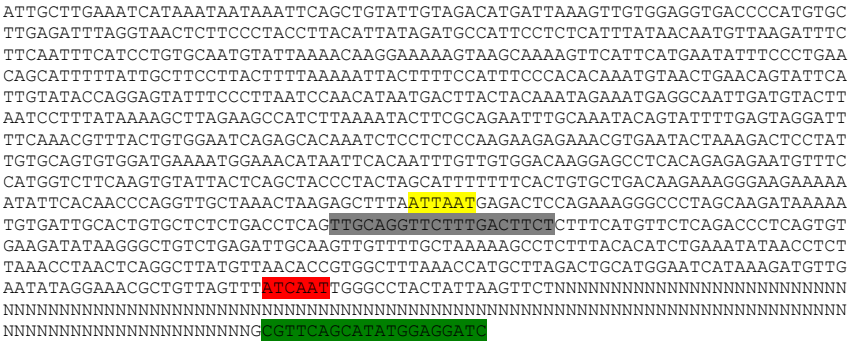


Figure 6. 5: Extra band of A375M from GAPS display gel.

DNA was extracted, cloned and sequenced. The *Ase1* restriction site is shown in yellow, the grey coloured box is the new primer. The polymorphism of the *Ase1* restriction site is shown in red. The CMK3LTR primer is highlighted in green colour.

A PCR experiment was set up with the above primers and 20 ng gDNA of eight cancer cells, as mentioned above for amplification of the expected 350bp product. 5 µl of PCR products were run on 1% gel agarose to confirm PCR was successful, and comparison to the 100bp marker was used to confirm the size of the product. The rest of the PCR product was split into three aliquots, one digested with the *Ase1* restriction enzyme, while the second aliquot was undigested by *Ase*1 restriction enzyme by using glycerol instead of enzyme, for three hours in a 37^{o}C water bath. The last aliquot was cloned and sequenced. This experiment included gDNA from SW480 as a positive control for enzyme activity. Our results observed the predicted two bands in A375M, compared with their absence from the other cancer cells including A375P (Figure 6.6). In addition, the sequence analysis showed observed a polymorphism in A375M (C instead of T), whereas A375P has a C in at this position (Figure 6.7). Thus, it is clear that the extra band was caused by a polymorphism in this *Ase1* restriction site in A375M only, and A375M was heterozygous for this RFLP.

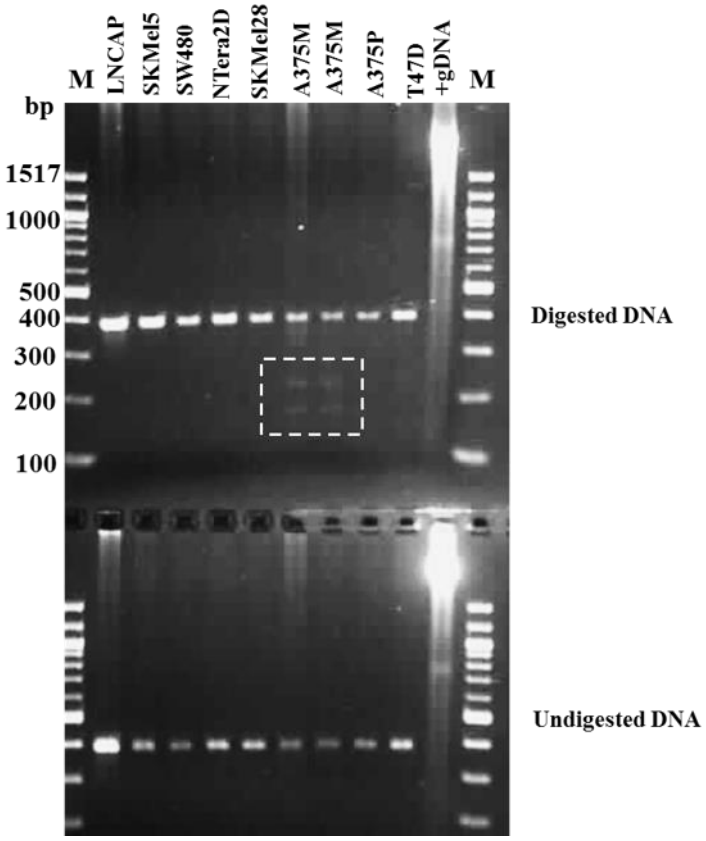


Figure 6. 6: polymorphism of restriction site in A375M cell.

Top panel shows the digested PCR products from eight cancer cells .Our data showed the extra bands in A375M cell (duplicate). The bottom part shows the undigested PCR products. This experiment included gDNA from SW480 as a positive (+) control.

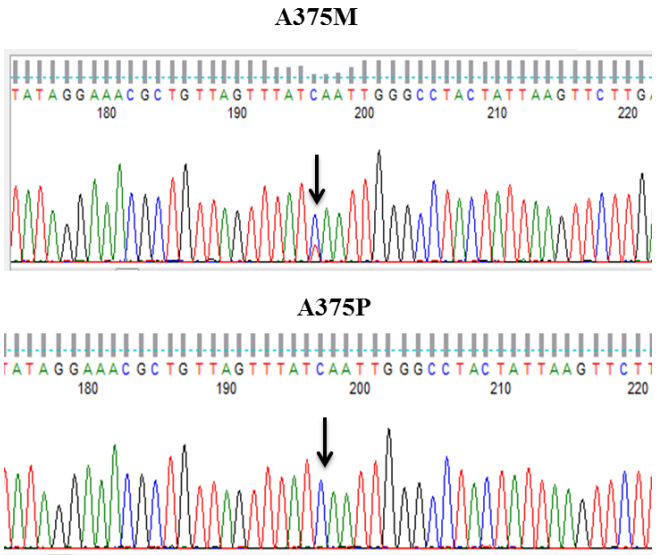


Figure 6. 7: Sequence data of polymorphism of Ase1 restriction enzyme.

Top panel - in sequencing reactions of A375M the observed two alleles (C/T) are indicated by a black arrow. Bottom panel – in the sequencing trace of A375P cell line only one allele (C) is indicated by a black arrow.

We concluded that the additional band observed in the A375M cell lines results from the creation of an *Ase*I site, in the heterozygous state, sufficiently close to a HERV-K provirus to generate a new amplicon. It is assumed that that this variant arose through a C>T change after the clonal cell lines had been derived.

6.2.4 NGS characterisation of HERV-K loci in cancer cell lines.

After making libraries of 8 duplicate gDNA cancer cells, the process of Ion torrent sequencing was carried out as described in the Material and Methods. Importantly, the quantification of the pooled barcoded libraries for top and bottom fractions was carried out individually using the Agilent High Sensitivity DNA Kit, according to the manufacturer's protocol. Libraries were pooled in equimolar proportions to prevent variation in sequencing coverage due to library concentration. In this way, subsequent sequence coverage variation could be attributed to differences in amplification efficiency rather than input pool variation.

Sequencing was the final stage of the process using an Ion torrent PGM instrument. Combining the GAPS assay with Ion torrent sequencing is proposed as a new method to discover novel insertions of HERV-K in human gDNA, as well as detecting segregating polymorphism. As the method is based on GAPS, which was effective in detecting novel polymorphic HERV-K proviruses in human and chimpanzee genomes, but enables analysis of distinct amplicons that co-migrate on agarose gels we expected its ability to detect insertions to far exceed that of the gel electrophoresis dependent variant.

Demultiplexed Ion Torrent data was analysed by Dr Richard Badge using the process illustrated in the pictured workflow (Figure 6.8) and used the Galaxy bioinformatics workbench (Macfarlane & Badge, 2015).
The workflow is available at [https://usegalaxy.org/u/rmb19/w/hervmacsanalysisnm] and was applied to each of the individually barcoded libraries. For the NGS-GAPS dataset this was 16 libraries in duplicate with 2 fractions (Top and Bottom) for a total of (in principle) 32 datasets. Unfortunately, one library from LNCAP was not amplified efficiently and was excluded from the pooling, resulting in 31 datasets.

In outline, the Galaxy pipeline involves checking the de-multiplexed and barcode clipped reads and quality scores were in appropriate format for mapping (fastqsanger, using the FastQ Groomer tool (Figure 6.8, A)) and then filtering the reformatted reads to remove any less than 42bp in length (the length of the ligated linker) (Figure 6.8, B). Reads were mapped to the hg38 assembly using BWA-MEM under the default settings, as captured in the workflow linked above (Figure 6.8, C). Due to the anticipated repetitive content of the reads, alignments were filtered to remove any with a mapping quality (MAPQ) of less than 30 (Figure 6.8, D).

The resulting BAM file was then sorted by chromosomal coordinate for facilitate downstream processing (Figure 6.8, E). To compensate for PCR bias due to differential amplification of some amplicons the Mark Duplicates tool of the Picard suite was used to remove PCR duplicates (identical read with identical alignments) (Figure 6.8, F). MACS analysis was then carried out on both the original and de-duplicated BAM files to assess the impact of duplicates on MACS peak calling (Figure 6.8, G). By subtracting the peaks called without PCR duplicates from the peaks called with duplicates, we

established if any additional peaks were recovered solely as a result of the inclusion of duplicates (Figure 6.8, I).

This comparison revealed that although across the analysis the number of peaks called from the duplicated set alone was a low and the majority of such calls had MACS scores of less than 100, and would be excluded by subsequent filtering. As a result, all analysis of the occurrence of HERV amplicons as reported as MACS peaks is conservative as only the de-duplicated BAM files were taken forward (Figure 6.8, H). It is possible some low scoring peaks represent true amplicons but their copy number in the sequenced library must have been lower than constitutional expectation (<~30 copies).

After processing with Galaxy pipeline the resulting MACs files were analysed to partition amplicons between known and novel HERV-K provirus loci. This initial dataset included around a thousand genomic, each with a different MACS score, related to the number reads mapping to the interval.

6.2.5 Analysis HERV-K loci identified by MACS mapping

The first step in this process was concatenation of the MACS file for each duplicate for each cell to make one file and their relabelling according to the cell origin and fraction (Top or Bottom) After this step, the data was filtered to remove any MACS peaks with a score of less than 100, as inspection of the data, using IGV, showed such peaks corresponded to single or small clusters of reads, rather than the 100s to 1000s of reads expected by coverage estimates. This excluded low confidence read clusters from the analysis, and again was a conservative measure.

The next step was to intersect the genomic locations of MACS peaks between pairs of cell line datasets, keeping the top and bottom fractions separate.

In this analysis, A375 was treated as one cell, so in total seven cell line dataset were analysed (A375, SKMel5, SKMel28, T47D, SW480 and NTera2D1), for both top and bottom fractions. The logic behind this analysis was that HERV-K proviral loci detected in more than two independently derived cell lines would most likely be constitutive insertions segregating in the population. The failure to detect a locus in any one cell line would be evidence for polymorphism of the insertion.

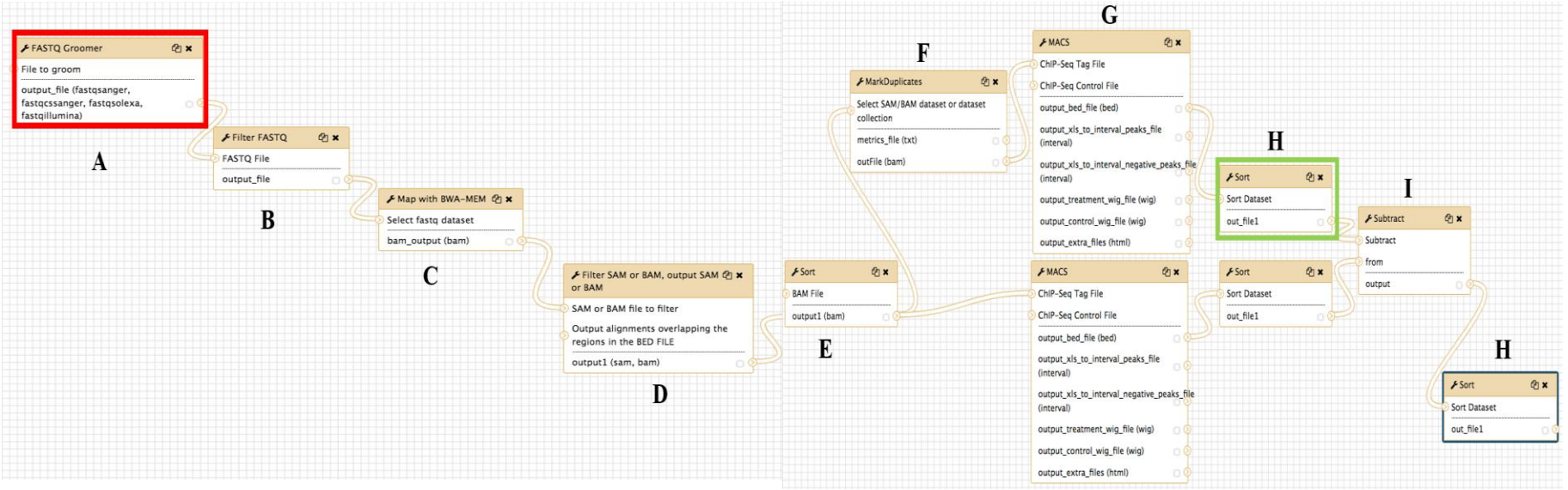


Figure 6. 8: Galaxy workflow for iontorrent analysis.

Input files (indicated by red box) at (**A**) were de-multiplexed and barcode clipped reads from the iontorrent instrument. Steps (**B**) to (**E**) were carried out in line and step (**G**), analysis with MACS, in parallel on mapped reads, with duplicates excluded (top) and included (bottom). Steps indicated by (**H**) involve sorting of the MACS output file by score, which is broadly related to sequence coverage. Step (**I**) involves subtraction of de-duplicated MACS results from the MACS analysis including duplicates to reveal peaks removed by de-duplication. Output file is boxed (green).

After excluding any loci that did not occur in at least two cell lines, the concatenate step was repeated again to collect putative HERV-K proviruses from all 7 cancer cell lines, separated into the top and bottom fractions also. This resulted in two data files, one for top fraction and another for the bottom fraction, collating all the high-confidence loci detected in our experiment. The hg38 Human Genome Assembly on the UCSC Genome Browser was used to analyse each locus in both of files and IGV software was used to visualise the read clusters as well as to verify the sequence of CMK3LTR primer and the linkered *Ase*1 site for reads. By combining the reported positions of 120 previously published proviruses (Subramanian *et al.*, 2011; Wildschutte *et al.*, 2016), translating these positions to the latest hg38 coordinates, and comparing the loci to our total dataset we observed that 32/120 proviral loci (Appendix 3) that we detected overlapped with previously published loci.

This data showed that NGS-GAPS on 7 cancer cell lines was effective in capturing a proportion of previously identified HERV-K proviruses. Interestingly, the NGS-GAPS method also detected 16 novel loci (Appendix 4), that were absent from the Subramanian / Wildeschutte combined dataset. As these loci correspond to annotated HERV-K LTRs within the hg38 assembly we suggest that these are previously undetected solo LTR/provirus polymorphisms, or proviruses that are erroneously represented by solo-LTRs in the genome assembly. We consider that this situation can be explained by positing that in our loci, proviruses were detected due to the proviral specificity of the GAPS method (Figure 6.9) which initiates amplification in the *env* gene, followed by the 3′ LTR. These same loci are represented in the reference genome as solo LTRs, which assisted their mapping, but suggests they may be proviruses. Thus, this is good evidence that the NGS-GAPS method isolates provirus sequences from across the genome. It also suggests that the number of intact proviruses in the human

genome is greatly underestimated, due to the assembly collapse of proviruses into solo-LTRs.

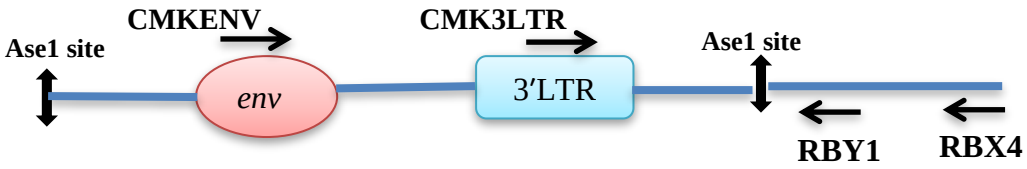


Figure 6. 9: Schematic diagram for GAPS assay.

HERV-K *env* and 3′LTR amplified by using CMKENV (Forward primers) and RBX4 (linker primers). Then, CMK3LTR (Forward primers) and RBY1 (linker primers) were used for nested PCR. *Ase1* was used as the restriction enzyme.

6.2.6 Detection of specific solo LTR/proviruses in each cancer cell line

The next part of the analysis was concerned with detection of proviral loci unique to each cell line. To achieve this location of the MACs peaks in each file were subtracted for each pairwise combination of cell lines to highlight loci only present in one of a pair of lines. Then, the name of the cell line was added to the location of the locus, and the selected loci concatenated, but with all top and bottom fractions kept separately. To exclude known proviruses the 120 published loci collated above (Subramanian *et al.*, 2011; Wildschutte *et al.*, 2016) were subtracted from the concatenated datasets. This resulted in a dataset of loci that had not been previously reported as harbouring a provirus and that were absent from at least one cell line in the pairwise comparisons.

Interestingly, this analysis captured 98 novel loci from 7 cancer cells (all of which were associated with a solo LTR in the reference genome, but by virtue of their amplification by 3′GAPS, are in fact likely to be proviruses. These 98 loci included 16 loci from intersection results, as illustrated in Appendix 5.

The distribution of these candidates was analysed across cell lines and 64 loci were found to be specific to one cell line (Table 6.1). The highest proportion of loci were found in NTera-2D1 (19/64) and the lowest proportion of loci were found in T47D (2/64) (Figure 6.10). Two loci were shared between two cell lines (LNCAP and NTera-2D1), and (LNCAP and SW480), respectively. The remaining 32 loci were found in all cell lines.

Table 6.1 Distribution of 64 novel loci in each specific cell lines.

Name of cells	Numbers of loci for specific cells
NTera 2D1	19
LNCAP	14
A375	13
SW480	6
SKMel5	5
SKMel28	5
T47D	2

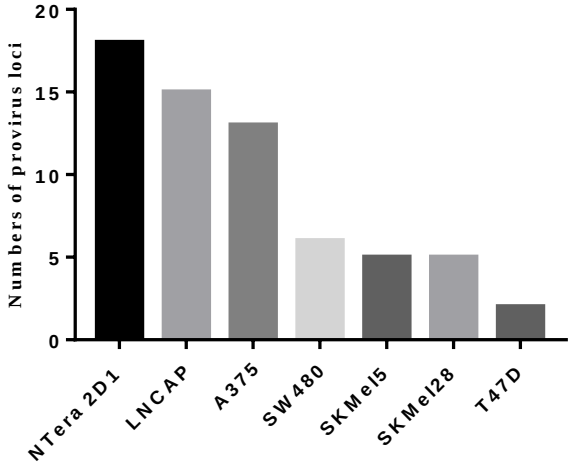


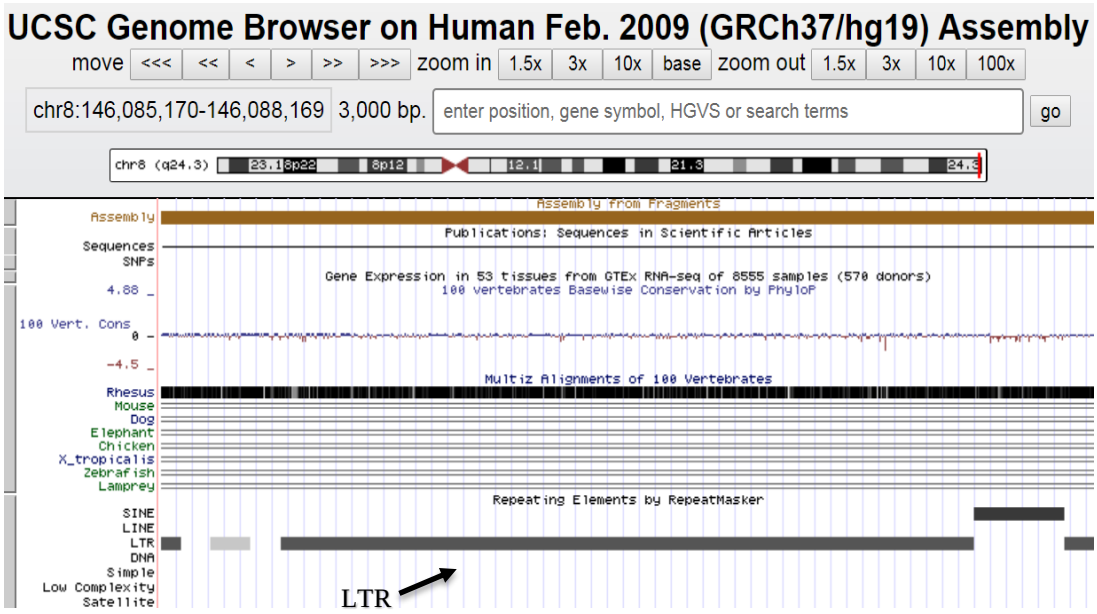
Figure 6. 10: Distribution of candidate solo LTR / provirus polymorphisms in 7 cancer cells lines. A bar chart shows the number of loci and the specific name of the cell line they occur in.

Previous research established that 120 loci were detected as novel proviruses from the literature (Subramanian *et al.*, 2011; Wildschutte *et al.*, 2016). However, our data showed only 9 loci (Appendix 6) detected in less than two cell lines were identical with loci reported in these papers.

This strongly suggests that by using proviral-specific amplification systems like GAPS we can reveal cryptic provirus / solo LTR polymorphisms, that are excluded by whole genome sequencing and assembly.

It can be seen from the Figure 6.11A that, for example, the provirus insertion at chr8:146085170-146088169 from Wildschutte *et al.* (2016) is a solo LTR in the reference genome, while they found solo LTR /provirus polymorphism. Likewise, our data showed novel loci, for example (chr2:192541525 -192542084). This locus is as a solo LTR in the reference genome while our result showed the presence of a provirus (6.11, B).

A



B

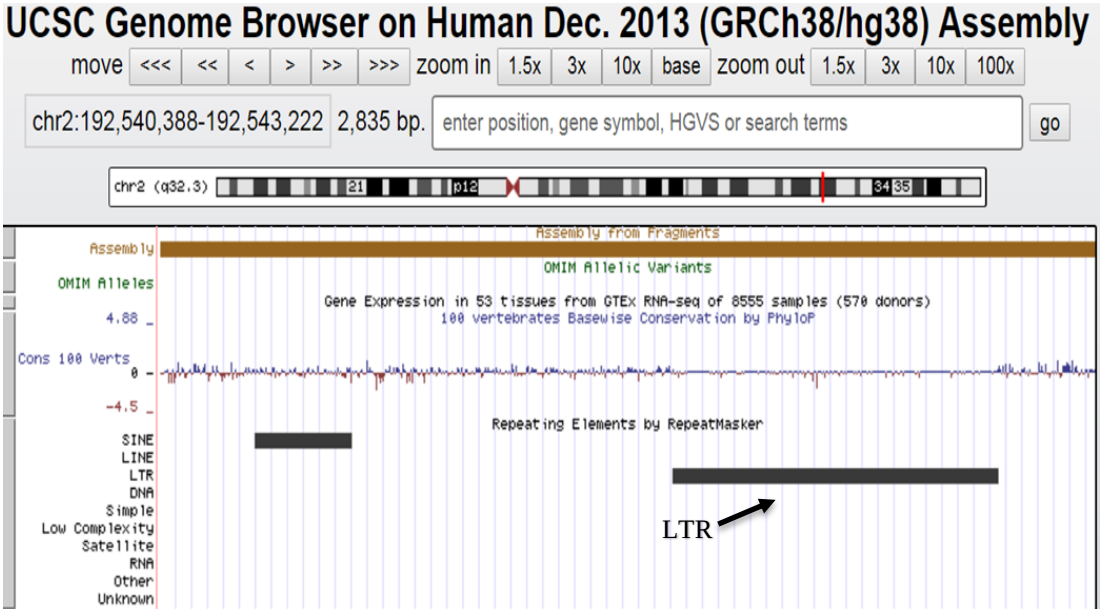


Figure 6. 11: Amplicon from LTR regions.

(A) This panel shows the HERV provirus at chr8:146085170-146088169 in (Wildschutte *et al.*, 2016) which observed the LTR region (**B**) The chr2:192541525-192542084location includes LTR region in reference genome compare to provirus detected by GAPS approach .The images were generated with the UCSC genome browser .

On other hand, many regions did not show HERV-K provirus because of reasons, such as mispriming, polymorphism in *Ase*1 site, PCR artefacts (Chimera) and re rearrangement by ligation of two loci together. All these artefactual amplicons were excluded from our data. The post MACS processing is summarized in Figure 6.12.

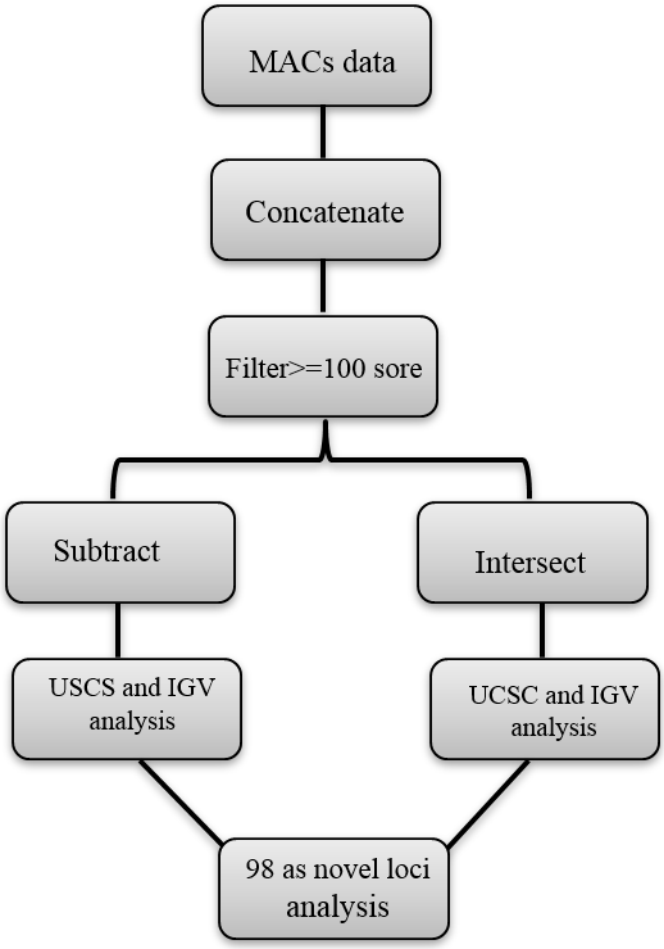


Figure 6. 12: Diagram of the process of analysis for HERV-K provirus loci.

All steps of analysis were done to detect solo LTR /provirus polymorphisms.

6.3 Discussion

Numerous previous studies have reported HERV-K provirus expression in different tissues associated with a variety of cancers including melanoma, breast, prostate, colon, and ovarian (Buscher *et al.*, 2006; Cegolon *et al.*, 2013; Chuong *et al.*, 2017). However, which HERVs are actively expressed (either at the RNA or protein level) in any given tissue is not well understood. For example, one study 1000 Genomes Project Consortium *et al.* (2015), analysed 20 different levels of HERV-K expression in abnormal and normal tissue, but they did not study which HERV-K loci showed presence / absence polymorphism in populations.

The GAPS approach, coupled with NGS, can be used to simultaneously detect proviral loci and their polymorphism status in populations. This approach was developed from the original GAPS protocol (Macfarlane & Badge, 2015). In the results presented here GAPS and Ion torrent NGS was applied to seven cancer cell lines of different origin (melanoma, breast, prostate, colon cancer and teratocarcinoma). The cell lines selected represent a limited population sample, but also included cell lines where HERV expression is well-established (A375, SKMel5, SKMel28, LNCAP, SW480 and T47D)

The most interesting finding was the detection of 98 candidate loci as novel proviral loci where previously these loci were annotated as solo LTRs. This result, while not validated by independent PCR assays suggests that proviruses are systematically excluded from assemblies and the census of HERV-K proviruses is much greater than represented in assembled genomes. A nineteen putative proviral loci reported were recorded in NTera2D1 while few novel loci (2/64) were amplified from T47D.

It can therefore be assumed that NTera2D1 cells have ability to produce VLP more than other cancer cells. Our data are in agreement with those obtained by Bieda *et al.* (2001), they found that the morphologically retrovirus-like particles (RVLP) are observed in different human teratocarcinoma cell lines. Moreover, they investigated three types of virions in teratocarcinoma cell lines: (1) the classical human teratocarcinoma-derived virus (HTDV) for example immature virus particles without spikes, which were present in all teratocarcinoma cell lines GH , Tera-1 cells , NCCIT , 2102Ep , and Tera-2 (2) immature virus particles with prominent spikes found in 2102Ep cells; and (3) free mature virus particles without surface spikes found in NCCIT cells (Bieda *et al.*, 2001).

In addition, to expression of VLP by teratocarcinoma cell lines, the different levels of HERV-Ks provirus may be associated with variation in numbers of provirus loci in different cancer cells. Thus, it could be the increased activation of HERV-K provirus in cells may have associated with numbers of provirus loci in comparison to solo LTR, but it is not related to a new insert HERV-K provirus in cancer cell. Our data corroborate the ideas of Bhardwaj *et al.* (2015) who noted that the majority of HERV-K RNA transcription originates from provirus loci in gDNA, and they found that (55.1%) of the expressed HML-2 sequences in teratocarcinoma cell lines is HERV-K *gag*. The increased number of intact or nearly intact proviruses could lead to increased HERV-K expression. However, according to our results, T47D included more solo LTRs than proviruses. Consistent with this observation T47D expressed low levels of the HERV-K *pol*, while no expression of *gag* protein was recorded (chapter 3, 4 respectively). These factors may explain the relatively good correlation between numbers of proviral loci and protein expression. Although, the mechanism of convert proviruses to solo LTR take a long time, it may happen quickly in some cells and as result, the cell will lose

expression of HERV-K proteins. Importantly, there is no evidence that this mechanism occurs in the opposite direction (solo LTR to provirus) (Macfarlane & Badge, 2015).

Taken together, these results give us a new view of on patterns of HERV-K proviral polymorphism in human populations. Some individuals may carry loci as proviruses while the others carry solo LTRs which have no ability to express protein. Furthermore, our data observed that two loci were shared between two cell lines despite our filtering approach (LNCAP and NTera-2D1), and (LNCAP and SW480). In summary, of the 98 novel putative proviral loci reported here, 32 loci were found in all cell lines, 64 were specific to one cell line and two were shared between two cell lines.

Our data are in agreement with the findings of Macfarlane & Badge (2015) who showed two HERV-K proviruses as novel insertions at 1p31.1a and 19 p12c in human and chimpanzees respectively. Previous studies have identified 120 proviral loci in human genomes by bioinformatic analysis (Subramanian *et al.*, 2011; Wildschutte *et al.*, 2016). One interesting aspect of our data is that only 9 of the loci were reported corresponded to these published proviruses. The GAPS approach was successful as it was able to confirm loci that were previously discovered, as well as reveal now loci.

From the intersection data, we discovered 16 novel candidate loci, detected in at least two cell lines, and an additional 32 loci that were identical with those detected in recently published analyses of existing sequence data (Subramanian *et al.*, 2011; Wildschutte *et al.*, 2016).

According to these data, we can infer that this method is suitable to detect proviral insertions (assuming the presence *env*-3′LTR) (Fig 6.10). The second advantage of using the GAPS method is that it allows detection of solo LTR vs provirus

polymorphisms. These advantages result from the GAPS method's property that it avoids the amplification of solo LTRs and SVA retrotransposons (Macfarlane & Badge, 2015).

Truthfully, it is difficult to analyse the expression of HERV-K proviruses and their contribution to human disease without understanding the polymorphic status of HERV-K provirus loci in the genome. Data from genome –wide associated studies (GWAS) observed that single nucleotide polymorphisms (SNP) in HERVs sequence was associated with a variety of diseases (Hindorff *et al.*, 2009).

There are still many provirus loci that have been not detected and the GAPS-NGS method presented here has limitations. One issue is that misspriming of CMK3LTR or RBY1 or deletion of some sequence of HERV-K in these primers may result in proviral loci not amplifying. Also, apparent insertional polymorphism may result from *AseI* site polymorphism as shown in figure (6.6).

This example showed an apparent polymorphism (extra band) in A375M was caused by a polymorphism in this *Ase1* restriction site (C instead of T) in A375M only. Another limitation of GAPS is that does not cover the entire genome because of the *Ase*I restriction site distribution. Only HERV proviruses within amplification range of *Ase*I sites can be detected, resulting in some loci being inaccessible.

Taken together, these results suggest that the novel proviral loci in human DNA may provide an important substrate for studying human evolution. Moreover, these data could update the catalogue of HERV-K provirus in the human population.

Although, The GAPS and Ion torrent approach did not reveal all HERV-K loci, but it could be used to detect many proviral loci. Further studies need to be carried out in order to validate, by locus specific PCR each candidate provirus locus, as some could represent false positives.

The most obvious finding to emerge from this study is that GAPS –NGS is a good approach to discover previously unreported provirus loci in gDNA. The 98 putative novel loci open a new door to understand the profile of HERV-K provirus and pattern of provirus polymorphism in human that may in turn associate with variation in levels of HERV-K gene expression.

7 Transcriptome characterisation of differentially cultured melanoma cell lines and melanocytes

7.1 Introduction

Some HERV-derived proteins have clear biological benefits in humans. For example, syncytin -1 is an essential protein derived from an ancestral HERV-W *env* gene. This protein causes fusion between the cell membranes of trophoblasts to form syncytiotrophoblasts. These structures are essential for normal development of the placenta during pregnancy (Schmitt *et al.*, 2013). However, HERV-K, the youngest of family of endogenized retroviruses is associated with a range of different human diseases, particularly cancer and especially melanoma (Buscher *et al.*, 2006; Serafino *et al.*, 2009).

Although the majority of HERV-K proviruses have lost their replicative function by mutation and recombination, some copies still have the ability to produce viral proteins and retrovirus viral like particles (RVLP). As a result many authors, observing that HERV-K mRNA, *gag, env*, np9 and *rec* were detected in melanoma cells compared to healthy individuals, have suggested that HERV-K plays a role in disease initiation or progression (Buscher *et al.*, 2005; Buscher *et al.*, 2006; Singh *et al.*, 2009).

Additionally, Serafino *et al.* (2009) demonstrated that a change in phenotype of melanoma cells in culture from adherent to non-adherent (suspension) was associated with increased HERV-K expression. Also, inhibition of HERV-K expression in this system led to reversion of the phenotypic change. These authors suggest that HERV-K expression directly drives metastatic transformation in melanoma, which results in a very poor clinical prognosis in melanoma patients.

In previous chapters, we described how serum starvation could induce phenotypic changes, specifically a change to a suspension growth form in one (SKMel5) of a panel of melanoma cell lines. This alteration in phenotype was associated with changes in

HERV-K expression, at the RNA and protein level, but not in ways consistent with the Serafino *et al.*, 2009 hypothesis.

Generally, NGS methods provide huge amounts of sequence information and applying NGS to cDNA derived from total cellular RNA allows scientists potentially to characterise the transcriptome in exhaustive detail. Sequencing steady-state RNA (RNA seq) is useful to study the transcriptome, as well as variation at the level of isoforms and novel transcription initiation sites (Oshlack *et al.*, 2010).

In this chapter, RNA seq was applied to melanoma cells of two cell lines that had been grown in two conditions (10% FBS and 1% FBS). Under these conditions one cell line (SKMel5) expresses a suspension growth phenotype, whereas the other (SKMel28) does not. Thus, comparing the transcriptome between cell lines and growth conditions will give insight into the changes in gene expression that are associated with the different phenotypes. As a control, we also characterised the transcriptome of normal primary melanocytes.

7.2 Results

7.2.1 Experimental design for RNA seq

The melanoma cell lines SKMel5 and SKMel28 were used for the RNA seq experiment. Three biological replicate cultures of these cells were grown in two culture conditions, 10%FBS and 1%FBS. At the same time melanocytes were cultured in 10% FBS (also with three biological replicates). For the SKMel5 cultures grown in 1% FBS, cell pellets were collected from both the adherent and suspended populations, as illustrated in Figure 7.1. All other cell pellets were collected from adherent cells. RNA was extracted and treated as described in the Methods and Materials.

The concentration and quality of RNA was evaluated by using an Agilent Bioanalyser and only high quality RNA samples from these cells (RIN numbers between 9 and 10) were selected. RNA was extracted from each of three biological replicates for each cell line, and once an RNA sample was selected, duplicate libraries were constructed from it, to give two technical replicates. Thus, each cell line and treatment was analysed in 6 libraries. As we sequenced libraries from SKMel28 grown in 10% and 1% FBS, adherent SKMel5 cells in 10% and 1% FBS, and suspended SKMel5 cells in 1% FBS and melanocytes, in all there were 36 different libraries, each distinguished by an illumina index (Table7.1).

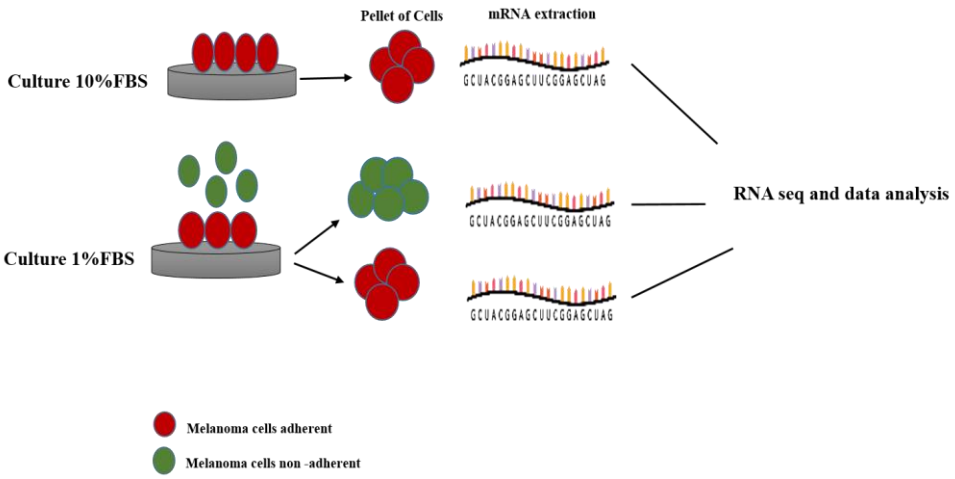


Figure 7. 1: Diagram of experiment setup for melanoma cells.

All cells were grown in two conditions 10% FBS and 1% FBS and the cell pellets collected for RNA seq. Adherent cells (red) were separated from suspension cells (green) by aspirating the media before trypsinisation.

Table 7.1 Identity, culture condition and phenotype of samples for the 36 RNA seq libraries prepared from melanoma cell lines and melanocytes.

Cell lines	Condition of culture (% FBS)	Phenotype	No. of libraries (3biological repeats , each with 2 technical replicates)
SKMel5	10%	Adherent	6
SKMel5	1%	Adherent	6*
SKMel5	1%	Suspension	6
SKMel28	10%	Adherent	6
SKMel28	1%	Adherent	6
Melanocytes	10%	Adherent	6

NB: *Due to concerns as to purity of the adherent population, and flow cell capacity, these libraries were not sequenced

7.2.2 Library preparation and sequencing

The libraries were prepared following the illumina TruSeq® Stranded Total RNA Sample Preparation Guide (using the low concentration sample protocol), as detailed in the Methods and Materials chapter. This ensured that all transcripts (with the exception of small RNAs and ribosomal RNAs which are specifically depleted), whether polyadenylated or not, were captured, and that the sequencing data generated was stranded.

Libraries were quantified and pooled in an equimolar design to be split across 4 NextSeq500 75bp High Output flow cells, and library QC and sequencing was carried out by the University of Nottingham NGS core service, Deep Seq. After sequencing, reads were split by library index, to pool reads from the same sample, biological and technical replicate, and the fastq files made available for download from DeepSeq's servers. The read data from the 30 sequenced libraries were split among 2 flow cells with four lanes each and yielded a total of ~1.5 billion indexed and Quality Control passed reads.

7.2.3 RNA seq data analysis

All intermediate data analysis was carried out by Dr Matthew Blades, of the University of Leicester Bioinformatics and Biostatistics Analysis Support Hub (BBASH). Briefly, paired end reads were mapped to the human genome reference (GRCh38) using the STAR alignment algorithm and sorted bam files loaded into the Partek Genomics suite (Ver 6.6) gene expression analysis module, for mRNA quantification and differential gene expression analysis. For 57.4% of paired reads both mapped within the exons of the same annotated gene, suggesting that, even without poly-A selection (total RNA was analyses) effective targeting to the protein coding transcriptome was achieved.

Dr Blades provided lists of significantly up or down regulated genes for selected pairwise comparisons, as detailed in table 7.2, filtered by a minimum fold change of 1.5 and at an FDR adjusted p-value of <0.05. This ensured that only statistically significant changes in gene expression were considered. These were subsequently filtered to minimum fold change of 2.0 and at an FDR adjusted p-value of <0.01; the top 100 upregulated and downregulated genes, for each comparison are presented in Appendix 7

Table 7.2 Numbers of differentially expressed genes for selected comparisons.

Comparison - Cell line (culture condition, phenotype)	No. DE Genes **(FC >1.5, FDR 0.05)**	No. DE Genes **(FC >2.0, FDR 0.01)**
Skmel28 (10% FBS, AD) vs Skmel28 (1% FBS, AD)	2078	92
Skmel5 (10% FBS, AD) vs Skmel5 (1% FBS, SUS)	7382	2545
Skmel5 (1% FBS, SUS) vs Skmel28 (1% FBS, AD)	12763	8400
Mel (10% FBS, AD) vs Skmel28 (10% FBS, AD)	13113	8852
Mel (10% FBS, AD) vs Skmel5 (10% FBS, AD)	14979	10756
Mel (10% FBS, AD) vs Skmel5 (1% FBS, SUS)	17069	12457

Columns indicate the pairwise comparison and number of differentially expressed genes between the pairs of cultures at default and more stringent (grey shading) fold change and FDR levels.

As this analysis revealed a very large number of differentially expressed genes, even between the same cell line simply grown in different concentrations of FBS, (e.g. Skmel28 (10% FBS, AD) vs Skmel28 (1% FBS, AD) -2078) we applied a more stringent fold change cut off and FDR level (>2 Fold Change, FDR < 0.01).

In most cases (Table 7.2, last column) this dramatically reduced the number of differentially expressed genes suggesting our original dataset has many genes with marginal changes in gene expression level and likely a high false positive rate. As a

result, we used only the more stringently filtered (FC >2, FDR < 0.01) data in subsequent analyses.

Numerical inspection revealed that the smallest difference in gene expression (Table 7.2) was seen between cultures of the SKMel28 cell line grown in 10% or 1%, which is not surprising as these cells grow robustly, even under serum starvation, and retain their adherent phenotype. The next smallest number of differentially expressed genes is seen between the Adherent culture of SKMel5 in 10% FBS and the Suspension phenotype population of the same cell line in 1% FBS (Skmel5 (10% FBS, AD) vs Skmel5 (1% FBS, SUS). As the main differences between these cultures is the altered phenotype induced by serum starvation, these genes are good candidates for being involved in the cellular and structural changes contributing to this change. However, with such a large number of significantly up and down regulated genes dissecting the responsible / responsive transcripts will require detailed functional analysis, which was constrained by the analysis time available.

In the comparisons of the transcriptome of normal human primary melanocytes and the three cultures, shown in the lower three rows of Table 7.2, it is not surprising that a very large number of gene are significantly differently expressed, although it is notable that the biggest difference in gene expression (by gene number) is between Melanocytes and SKMel5 cells expressing the suspension phenotype, in 1% FBS. This finding emphasizes how dramatically different, at the transcriptome level, cultured melanoma cells are, when compared to their originator cell type.

7.2.4 Gene Ontology and functional clustering analysis

All gene lists were input to the DAVID Functional Annotation Tool website to detect statistically enriched Gene Ontology terms, functional annotations and pathway membership within the lists of differentially expressed genes. For each comparison in table 7.2 the gene lists were sorted by fold change, and the 100 most up-regulated genes, and 100 most down regulated genes, in the first member of the comparison (e.g. SKMel28 adherent cells in 10% FBS in the "Skmel28 (10% FBS, AD) vs Skmel28 (1% FBS, AD)" comparison were extracted. This is necessarily an arbitrary sampling strategy but should capture those genes whose expression is most greatly affected by cell origin, culture condition and phenotype.

The DAVID Bioinformatics Resources Website 6.8 (https://david.ncifcrf.gov/ , updated to version 6.8 in 2018) was used to analyses the top 100 most up and down regulated genes from each comparison, using the David Functional Clustering (DFAC) metanalysis. This tool combines enrichment of terms associated with unique gene identifiers (we used HGNC gene symbols as a non-redundant, curated set of identifiers) from the input list versus the same set of terms from the entire human gene corpus. The approach uses a modified Fisher's Exact test (with Benjamini Hochberg correction for multiple testing) to detect statistically significant enrichment for particular terms. The advantage of this analysis is that it encompasses annotation terms from a wide variety of sources (functional annotation, Gene Ontology and pathway membership). For these analyse, we restricted the annotation sources to Uniprot functional annotation (COG Ontology, Uniprot_Keywords, Uniprot_sequence_features) Gene Ontology (Biological Process, Molecular Function, Cellular Component) and Pathway membership (KEGG_pathway).

7.2.5 DAVID Functional Annotation Clustering (DFAC) in serum starvation

The differential gene expression comparison that generated the smallest number of differentially expressed genes was that between SKMel28 adherent cells grown in 1% FBS versus 10%.

As these cells are derived from the same population of cells we might expect few differences, except these connected with the 10-fold lower serum components.

Interestingly, on analysis of all 92 differentially expressed genes, a strong functional analysis cluster (Enrichment score 6.27), illustrated in Table 7.3, was detected. Despite the small number of genes in the list the cluster is dominated by terms associated with cholesterol and general steroid biosynthesis. This result would seem to suggest that the main response of SKMel28 cells to serum starvation is up-regulation of genes involved in cholesterol biosynthesis, which is consistent with the high concentration of cholesterol in Fetal Bovine Serum (~300ug / ml in some preparations [URL https://fscimage.fishersci.com/cmsassets/downloads/segment/Scientific/pdf/Cell_Culture/Application_Notes/Lipids_Media.pdf)]. This also demonstrates that applying Functional Annotation Clustering is an effective approach to find biologically relevant information in our comparisons.

Table 7.3 Differential expressed (up and down) genes comparing SKMEL28 in 1% FBS versus SKMEL28 in 10% FBS.

Analyzed using DAVID Functional Annotation Clustering. The Category column specifies which data resource the enriched term originated from (GOTERM = Gene Ontology, UP_ = UniProt, KEGG_ = Kyoto Encyclopedia of Genes and Genomes). The Term column lists the enriched term itself, the Count column is the number of genes associated with the enriched term. The Genes column list the identifiers of the genes showing the enriched term. The Fold Enrichment (F.E) column shows the number of times the term is enriched in the gene set, above background expectation and the p-value column lists the adjusted (Benjamini-Hochberg) p-value under a null hypothesis of no specific enrichment.

Cluster1 Enrichment Score: 6.27

Category	Term	Count	Genes	Fold Enrichment F.E.	p-value (BH)
GOTERM_BP_DIRECT	GO:0006695~cholesterol biosynthetic process	11	TM7SF2, MSMO1, MVD, SQLE, HMGCR, DHCR7, INSIG1, FDPS, HMGCS1, LSS, IDI1	118.56	1.87E-16
UP_KEYWORDS	Steroid biosynthesis	9	TM7SF2, MSMO1, MVD, HMGCR, DHCR7, FDPS, HMGCS1, LSS, IDI1	110.26	1.07E-12
UP_KEYWORDS	Sterol biosynthesis	8	TM7SF2, MSMO1, MVD, HMGCR, DHCR7, FDPS, HMGCS1, IDI1	137.21	5.77E-12
UP_KEYWORDS	Sterol metabolism	9	TM7SF2, MSMO1, MVD, HMGCR, DHCR7, INSIG1, FDPS, HMGCS1, IDI1	57.60	9.47E-11
UP_KEYWORDS	Cholesterol biosynthesis	7	TM7SF2, MVD, HMGCR, DHCR7, FDPS, HMGCS1, IDI1	157.97	8.24E-11
UP_KEYWORDS	Steroid metabolism	9	TM7SF2, MSMO1, MVD, HMGCR, DHCR7, INSIG1, FDPS, HMGCS1, IDI1	44.87	4.48E-10
UP_KEYWORDS	Cholesterol metabolism	8	TM7SF2, MVD, HMGCR, DHCR7, INSIG1, FDPS, HMGCS1, IDI1	60.18	1.00E-09

UP_KEYWORDS	Lipid biosynthesis	9	TM7SF2, MSMO1, MVD, HMGCR, DHCR7, FDPS, HMGCS1, LSS, IDI1	24.74	3.88E-08
KEGG_PATHWAY	hsa01130:Biosynthesis of antibiotics	11	TM7SF2, MSMO1, MVD, SQLE, HMGCR, FDPS, HMGCS1, LSS, PCK2, ACAT2, IDI1	11.95	4.25E-07
KEGG_PATHWAY	hsa00900:Terpenoid backbone biosynthesis	6	MVD, HMGCR, FDPS, HMGCS1, ACAT2, IDI1	62.82	6.70E-07
GOTERM_BP_DIRECT	GO:0008299~isoprenoid biosynthetic process	5	MVD, HMGCR, FDPS, HMGCS1, IDI1	146.27	3.26E-06
UP_KEYWORDS	Lipid metabolism	10	TM7SF2, MSMO1, MVD, HMGCR, DHCR7, INSIG1, FDPS, HMGCS1, LSS, IDI1	9.90	7.61E-06
KEGG_PATHWAY	hsa00100:Steroid biosynthesis	5	TM7SF2, MSMO1, SQLE, DHCR7, LSS	57.58	2.28E-05
KEGG_PATHWAY	hsa01100:Metabolic pathways	15	TM7SF2, GCNT3, MSMO1, MVD, HMGCR, HMGCS1, FDPS, LSS, ACAT2, PCK2, NMRK2, SQLE, DHCR7, IDI1, AMD1	2.81	0.001857
GOTERM_BP_DIRECT	GO:0016126~sterol biosynthetic process	3	TM7SF2, MSMO1, SQLE	122.87	0.019489
UP_KEYWORDS	Endoplasmic reticulum	9	TM7SF2, TMEM97, MSMO1, SQLE, HMGCR, DHCR7, INSIG1, LSS, ATP10D	3.62	0.035117

7.3 Differential gene expression in Melanocytes and melanoma cell lines

Comparisons of the Melanocyte transcriptome with that of melanoma cell lines with varying phenotype under different culture conditions yielded the largest number of differentially expressed genes (between 8852 and 12457 genes Table 7.2). This large number, relative to the number of genes in the human genome (25-30,000) means that we lack statistical power to detect enrichment of functional annotation within the complete dataset. Restricting the analysis to the top 100 up and down regulated genes makes detecting enrichment possible, but is biologically an arbitrary approach – if highly differentially expressed genes do not happen to fall into this range then functional enrichment will not be detectable. Despite this caveat, we executed the DFAC analysis on each gene list from Table 7.2, for the 100 most up and down regulated genes. All comparison with lists of down regulated genes yielded no significant functional annotation clusters. Three of the comparisons DFAC analyses of up regulated genes (i.e. genes that were more highly expressed in melanocytes than the comparator cell) did show enrichment, and the most significant functional annotation clusters are illustrated in Tables 7.4, 7.5 and 7.6. However, while the illustrated annotation clusters are statistically significant, there are few individual terms whose enrichment is significant after Benjamini-Hochberg multiple testing correction. Ultimately the observed enrichment of extracellular matrix / space, cellular adhesion and protein secretion / glycosylation associated GO terms and functional annotation in these clusters, is consistent with melanocytes existing in an integrated tissue (skin), as opposed to culture adapted malignant cells growing in pure culture. However, the enrichment is modest and due to the sampling strategy may be insensitive or prone to false positive clusters, so interpretation must be cautious.

7.4 Differential gene expression in melanoma cell lines with altered phenotypes

The primary motivation to analyse the transcriptome of the panel of cell lines studied here was to investigate whether the changes in phenotype we observed upon serum starvation of one melanoma cell line (SKMel5) was associated with changes in patterns of gene expression that might explain this phenotypic change. Comparing the transcriptome of SKMel5 suspension cell line with either normal primary human melanocytes, or the original SKMel5 population that had been maintained in 10% FBS yielded large numbers of differentially expressed genes (12457 and 2455 respectively) limiting our statistical power to detect functional enrichment. Nevertheless, the analysis procedure did yield one moderately enriched cluster (Enrichment score 1.23, illustrated in table 7.7), among the top 100 up regulated genes in SKMEL5 in 10% FBS (AD) versus SKMEL5 in 1%FBS (SUS). These are genes expressed at higher levels in SKMel5 cells growing adherently in 10% FBS, than in SKMel5 cells growing as a suspension culture in 1%, and include genes associated with the GO term "GO:0005576~extracellular region" which may be biologically relevant. This term occurs in two overlapping groups of genes, and each group is individually significant ($p < 0.05$), but non-significant after multiple testing correction.

A notable feature of this comparisons differentially expressed gene list (SKMEL5 in 10% FBS (AD) versus SKMEL5 in 1%FBS (SUS)), is that 89 of the top 100 up regulated genes (i.e. implicitly down – regulated in the suspension (SUS) cells) are ribosomal or nucleolar small RNA genes, tRNAs or pseudogenes. It is possible that this reflects a technical failure of ribosomal depletion, but this would not account for the enrichment of pseudogenes. An intriguing possibility is that this results from a reduction in spurious / excessive transcription of small RNA genes and pseudogenes, or

perhaps more active RNA recycling due to cellular metabolic stress in starvation condition. This could cause lower steady state levels of these sequences, but it is necessary to exclude non-biological explanations first.

Table 7.4 Top 100 upregulated genes comparing Melanocytes and SKMEL28 in 10%FBS.

DAVID Functional Annotation Clustering. The Category column specifies which data resource the enriched term originated from (GOTERM = Gene Ontology, UP_ = UniProt, KEGG_ = Kyoto Encyclopedia of Genes and Genomes). The Term column lists the enriched term itself, the Count column is the number of genes associated with the enriched term. The Fold Enrichment (F.E) column shows the number of times the term is enriched in the gene set, above background expectation and the p-value column lists the adjusted (Benjamini-Hochberg) p-value under a null hypothesis of no specific enrichment**.**

Annotation Cluster 1. Enrichment Score: 2.47

Category	Term	Count	p-value	F. E.	p-value (BH)
UP_KEYWORDS	Signal	29	5.80E-04	1.86	0.080622936
UP_SEQ_FEATURE	signal peptide	25	8.11E-04	1.97	0.212198316
UP_KEYWORDS	Glycoprotein	30	0.001106236	1.76	0.077111177
UP_KEYWORDS	Disulfide bond	24	0.002379128	1.87	0.108748135
UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	27	0.004556141	1.68	0.488947554
UP_SEQ_FEATURE	disulfide bond	20	0.009594366	1.81	0.611237821
GOTERM_CC_DIRECT	GO:0005615~extracellular space	12	0.011913699	2.32	0.444160982
UP_KEYWORDS	Secreted	14	0.027002251	1.90	0.432788324

Annotation Cluster 3 Enrichment Score: 1.7590910903597476

Category	Term	Count	p-value	F. E.	p-value (BH)
GOTERM_BP_DIRECT	GO:0030198~extracellular matrix organization	7	1.09E-04	9.087	0.045755003
GOTERM_MF_DIRECT	GO:0005201~extracellular matrix structural constituent	5	1.28E-04	19.08	0.0179978

Table 7.5 Top 100 upregulated genes comparing Melanocytes and SKMEL5 in 10%FBS.

DAVID Functional Annotation Clustering. Annotation. The Category column specifies which data resource the enriched term originated from (GOTERM = Gene Ontology, UP_ = UniProt, KEGG_ = Kyoto Encyclopedia of Genes and Genomes). The Term column lists the enriched term itself, the Count column is the number of genes associated with the enriched term. The Fold Enrichment (F.E) column shows the number of times the term is enriched in the gene set, above background expectation and the p-value column lists the adjusted (Benjamini-Hochberg) p-value under a null hypothesis of no specific enrichment. Cluster 1Enrichment Score: 2.64

Category	Term	Count	p-value	F. E.	p-value (BH)
UP_KEYWORDS	Calcium	14	2.72E-05	4.11	0.004508181
UP_KEYWORDS	Cell adhesion	10	9.03E-05	5.37	0.007467059
GOTERM_BP_DIRECT	GO:0034332~adherens junction organization	4	4.67E-04	25.93	0.236881447
GOTERM_BP_DIRECT	GO:0007156~homophilic cell adhesion via plasma membrane adhesion molecules	6	4.79E-04	9.11	0.129491845
UP_SEQ_FEATURE	domain:Cadherin 5	5	7.15E-04	12.25	0.112889558
UP_SEQ_FEATURE	domain:Cadherin 3	5	9.10E-04	11.48	0.096689069
UP_SEQ_FEATURE	domain:Cadherin 4	5	9.10E-04	11.48	0.096689069
UP_SEQ_FEATURE	domain:Cadherin 2	5	0.001004445	11.18	0.080720002
UP_SEQ_FEATURE	domain:Cadherin 1	5	0.001004445	11.18	0.080720002
GOTERM_BP_DIRECT	GO:0007155~cell adhesion	8	0.002727335	4.18	0.409682483
GOTERM_CC_DIRECT	GO:0005913~cell-cell adherens junction	6	0.011240661	4.40	0.336829678
KEGG_PATHWAY	hsa04514:Cell adhesion molecules (CAMs)	4	0.015632309	7.21	0.683417715
GOTERM_MF_DIRECT	GO:0005509~calcium ion binding	8	0.030726722	2.62	0.754483442
UP_KEYWORDS	Cleavage on pair of basic residues	4	0.103741938	3.50	0.837673641
GOTERM_CC_DIRECT	GO:0016020~membrane	10	0.577288589	1.08	0.997165839

Table 7.6 Top 100 up regulated genes comparing Melanocytes and SKMEL5 in 1%FBS.

DAVID Functional Annotation Clustering. Annotation. The Category column specifies which data resource the enriched term originated from (GOTERM = Gene Ontology, UP_ = UniProt, KEGG_ = Kyoto Encyclopedia of Genes and Genomes). The Term column lists the enriched term itself, the Count column is the number of genes associated with the enriched term. The Fold Enrichment (F.E) column shows the number of times the term is enriched in the gene set, above background expectation and the p-value column lists the adjusted (Benjamini-Hochberg) p-value under a null hypothesis of no specific enrichment.

Cluster 1. Enrichment Score: 2.18

Category	Term	Count	p-value	F. E.	p-value (BH)
UP_SEQ_FEATURE	signal peptide	26	9.17E-05	2.195760336	0.024814783
GOTERM_CC_FAT	GO:0044421~extracellular region part	26	7.68E-04	1.837661895	0.135862754
GOTERM_CC_FAT	GO:0005576~extracellular region	28	0.002029417	1.660099829	0.175510174
UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	26	0.003422239	1.735241872	0.171269534
UP_KEYWORDS	Glycoprotein	27	0.003789336	1.695863546	0.187422363
GOTERM_CC_FAT	GO:0005615~extracellular space	13	0.004250813	2.467608321	0.236461705
UP_KEYWORDS	Secreted	15	0.006608317	2.182039864	0.238023977
GOTERM_CC_DIRECT	GO:0005615~extracellular space	12	0.00959979	2.38752784	0.615176452
UP_KEYWORDS	Disulfide bond	18	0.073505036	1.498325568	0.679623823
UP_SEQ_FEATURE	disulfide bond	15	0.131030703	1.453089466	0.978683228
GOTERM_BP_FAT	GO:0002682~regulation of immune system process	8	0.201433232	1.63833608	0.931387159

Table 7.7 Top 100 up regulated genes comparing SKMEL5 in 10% FBS (AD) and SKMEL5 in 1%FBS (SUS).

DAVID Functional Annotation Clustering. Annotation. The Category column specifies which data resource the enriched term originated from (GOTERM = Gene Ontology, UP_ = UniProt, KEGG_ = Kyoto Encyclopedia of Genes and Genomes). The Term column lists the enriched term itself, the Count column is the number of genes associated with the enriched term. The Fold Enrichment column shows the number of times the term is enriched in the gene set, above background expectation and the p-value column lists the adjusted (Benjamini-Hochberg) p-value under a null hypothesis of no specific enrichment.

Cluster 1. Enrichment Score: 1.23

Category	Term	Count	p-value	Genes	F. E.	p-value (B/H)
GOTERM_CC_DIRECT	GO:0005576~extracellular region	5	0.008250799	MTRNR2L12, MTRNR2L2, MTRNR2L1, FIBIN, CD200R1	5.15	0.152700042
UP_KEYWORDS	Secreted	5	0.010859919	MTRNR2L12, MTRNR2L2, MTRNR2L1, FIBIN, CD200R1	4.76	0.302559102
GOTERM_CC_FAT	GO:0005576~extracellular region	7	0.033814412	KRT81, MTRNR2L12, MTRNR2L2, MTRNR2L1, FIBIN, KRT86, CD200R1	2.20	0.73856625
UP_KEYWORDS	Cytoplasm	4	0.425945806	SGCG, MTRNR2L12, MTRNR2L2, MTRNR2L1	1.55	0.898683381
GOTERM_CC_DIRECT	GO:0005737~cytoplasm	4	0.580584319	SGCG, MTRNR2L12, MTRNR2L2, MTRNR2L1	1.27	0.996950015

7.5 Discussion

Consideration of the results of differential gene expression comparisons leads us to conclude that the levels of differentiation in gene expression patterns do reflect biological changes, resulting from different cellular genotypes, cell type, culture conditions, and growth phenotype. However, there are limitations to the analysis and its interpretation that could be addressed by further work. It should be noted that analyses of the RNAseq datasets reported in this thesis are necessarily preliminary. Due to circumstances beyond our control (library pooling issues, core facility shutdown, data analysis bottlenecks) the differential gene expression analysis results were available only shortly before the submission deadline. We have attempted to identify the main trends in gene expression variation between selected cell populations, but detailed analysis was made impossible by time constraints.

A primary limitation that should be addressed is the large numbers of differentially expressed genes. This may reflect biological reality, but it is also possible that some genes appear to be differentially expressed as a result of biological and technical noise. In this preliminary analysis, biological and technical replicates were pooled, and although all datasets passed internal QC, it is possible that some unusual replicates are significantly affecting the analysis. We assume that quantifying technical variation will enable us to set thresholds for variation induced in library preparation and sequencing from identical RNA samples. If this is the case then comparing differential gene expression between biological replicates of a cell line / culture condition will allow us to restrict analysis to genes that are consistently up or down regulated across independent cultures. This should have the effect of reducing the numbers of differentially expressed genes, while enriching for those contributing to variation in the transcriptome.

Finally, one set of RNAseq libraries was not sequenced (RNA derived from adherent SKMel5 cells growing in 1% FBS). The primary reason for this was the fact that we could not be sure that the adherent and suspension populations were separable, and also lacked capacity on our planned number of flow cells. Clearly as we can detect many differentially expressed genes between SKMel5 adherent and suspension cells, grown separately, if these differences were validated, we could sequence these other libraries to investigate patterns of gene expression within the SKMel5 1% FBS Adherent cultures. Potentially this could enable us to address which of the changes in gene expression are specifically related to serum starvation and which are related to the change in phenotype.

8 General discussion

8.1 Introduction

Tranposable elements (TE) are endogenous components of the genomes of eukaryotes, constitute around 45% of the human genome DNA. Retrotransposons are the major class within these transposable elements. Recent studies have revealed that LTR elements include human endogenous retroviruses (HERV), which constitute about 8% of human DNA. HERVs are derived from sequences integrated into germ cells during exogenous retrovirus infection, up to 25 million years ago (Lander *et al.*, 2001). However, most copies of HERVs are defective in multiple ways. For example, recombination occurs between the two LTRs to form a single LTR (solo LTR). The HERV-K family of Class II HERVs is the youngest family and likely still has significant biological activity, because many elements have intact ORFs. As a result, some sequences of this family can produce viral particles (Kassiotis, 2014; Suntsova *et al.*, 2015).

HERV-K activity may also be involved in a variety of diseases. Many researchers have proposed a causative role for HERVs in autoimmune diseases, for example, rheumatoid arthritis (RA) (Serafino *et al.*, 2009) and multiple sclerosis (MS) (Cegolon *et al.*, 2013). Moreover, HERV-K may play a role in cancer, especially in human melanoma (Muster *et al.*, 2003) and is correlated with genome-wide hypomethylation conditions (Stengel *et al.*, 2010).

8.2 Quantification of HERV-K *pol* gene expression in cancer cells, under two conditions

To investigate the correlation between HERV-K and cancer or its development, different cancer cell lines were used: melanoma cell lines (A375P, A375M, SKMel5 and SKMel28 and other cancer cells: LNCAP, SW480, T47D. These cell lines were grown in media containing 10% FBS as the normal culture condition, while 1% FBS media was used to induce starvation conditions. Our experiments showed that phenotype changes occurred only in the SKMel5 cell line growing in starvation conditions (1% FBS), and this change was not typical of melanoma cell lines. The phenotype observed was clusters of round-shaped cells, detached from the adherent monolayer, growing actively in suspension specifically in 1% FBS, as compared with when 10% FBS was included in the media. That these cells were viable, but expressing a different phenotype, rather than dead or dying cells that had lost the ability to adhere to the culture vessel surface, was confirmed by colony formation and viability assays. Ultimately, only one melanoma cell showed any phenotypic change (SKMel5). It should be noted that the A375 lines showed slightly different phenotypes, but did not change as much as SkMel5. The phenotype was different to that observed by Serafino *et al*., (2009) paper, and we do not know which A375 Serafino used as their negative control.

HERV-K *pol* gene expression in all melanoma cell lines and other cancer cell lines was investigated in the two culture conditions, to investigate the response of the HERV-K 5′ LTR promoter to serum starvation. The *pol* gene was used as a proxy for the activity of the HERV-K 5′ LTR as although HERV proviruses generate a range of transcripts, the two main transcripts, encoding the *gag-pro-pol* polyprotein and the *env* proteins share the same promoter.

We observed that expression of *pol* was up regulated in A375P under 1% FBS compared with 10% FBS culture, with a 5.6 fold change, while for other melanoma cell lines there was no statistically significant difference in expression between conditions. In addition, cell lines derived from other cancer types (prostate, colon, breast) grown in 1% FBS showed upregulation of *pol* expression, especially the SW480 metastatic colon cancer cell line.

With regards to SKMel5, which showed a change in phenotype in serum starvation conditions, expression of *pol* decreased according to condition and phenotype, but a statistically significant difference in expression *pol* was not observed. It is possible that the expression of HERV-K is not up regulated under starvation conditions that induce the modification of cellular phenotype from adherent to suspension, and which is not specifically associated with increased or reduced expression, in our hands. This is in contrast to the study of Serafino *et al.* (2009) who observed HERV-K overexpression when melanoma cell lines (TVM-A12 and M14) transitioned from adherent to suspension in 1% FBS culture. These authors reported that A375 was negative with respect to this change in morphology, although in their paper they did not report A375 *pol* expression, nor specify which of the two A375 (P or M) sub-lines they used. This observation was also reported in a later publication by the same group (Argaw-Denboba *et al*. 2017), and the authors did utilize A375 as a negative control.

Also, there was no difference in the level of HERV-K mRNA of *pol* expression between melanoma cell lines and normal primary melanocytes. These results raise the possibility that HERV-K *pol* expression is not related with the development stages of melanoma, in contrast to that of *env*, which is widely reported as an informative biomarker of melanoma (Nadeau *et al.*, 2015). This observation is interesting as it

suggests that association of HERV-K *env* expression is specific to the *env* transcript, rather than reflecting general upregulation of the HERV-K promoter.

8.3 Detection of *gag* and *env* proteins in cancer cells under two conditions

Our results showed that *gag* protein expression was detected in all melanoma cells, a prostate cancer cell line (LNCAP), a colon cancer cell line (SW480), but not a breast cancer cell line (T47D), in normal culture conditions (10% FBS). The *gag* protein is produced, along with *pol* and *pro*, as a polyprotein and released by cleavage of the polyprotein, by the protease *pro*. Thus, it is expected that *pol* RNA levels and *gag* protein levels should be closely correlated. In contrast, as the *env* protein is translated from an alternatively spliced transcript (that shares the HERV-K 5′ LTR) it is possible, through higher rates of splicing and translation, that *env* protein levels could be independent of *pro* or *gag* levels.

We observed that in 1% FBS culture conditions all cancer cell lines significantly down regulated *gag* expression. Surprisingly, HERV-K *env* protein expression increased in SKMel5 adherent cells in 1% FBS, compared to the same culture of SKMel5 cells in 1%, showing their suspension phenotype. This suggests that, consistent with previous reports, that *env* up regulation is associated with a change in cellular phenotype, but that this is transient, and is not associated with maintain the suspension phenotype. It may be the case that previous studies combined the different patterns of *env* expression in the two cellular populations (adherent and suspension) thus associating elevated *env* expression with altered cellular morphology. As we were able to separate the two cell populations, and show divergent patterns of *env* expression, this indicates that *env* expression is more dynamic than previously appreciated. Importantly these dynamic

changes in *env* protein expression occur in the context of *gag* down regulation, supporting the idea that although apparently driven by the same promoter, the main outputs of HERV transcriptional activity, *gag* and *env* protein levels are largely de-coupled.

If *env* is directly responsible for the alterations in cellular phenotype that are thought to be relevant to melanoma disease progression, this result suggests that these effects are somewhat independent of HERV-K provirus activity.

Another observation from these experiments was that HERV-K *env* protein expression increased in T47D cells in 1% FBS, compared with those grown in 10% FBS, in contrast to reduced *gag* expression. This finding supports the decoupling of HERV-K *gag* and *env* expression, but also accords with the association of *env* expression with breast cancer development (Johanning *et al.*, 2017).

In summary, serum stress conditions may contribute to increased expression of HERV-K proteins in one melanoma cell line, but, it is not a phenomenon seen for all melanoma cell lines, and the suspension morphology is not dependent upon increased protein expression. In contrast, Serafino *et al*. (2009) demonstrated a positive correlation between their observed change in phenotype and increased expression of HERV-K *env* protein, as well as demonstrating reversion of the phenotype when *env* expression was specifically inhibited. Taken together these results suggest that HERV-K *env* does have a role in phenotypic changes in some melanoma cell lines, but that this role may be transient, and modified by other factors, rather than being a property of melanoma exclusively.

We also investigated the expression of these HERV-K proteins in normal primary, melanocytes, as the cell type responsible for melanoma. We observed significantly lower, but measurable, expression of *gag* protein, compared with melanoma cells, consistent with higher HERV-K promoter activity in melanoma.

However, *env* protein expression was not detected in melanocytes, again supporting decoupling of HERV expression, at least at the translational level despite the HERV-K promoter being active in these cells (as judged by *gag* expression). These data are in agreement with that of Buscher *et al*. (2005), who found that *gag* and *env* HERV-K proteins were expressed in melanoma cells, but normal neonatal melanocytes did not express *env* protein.

Taken together these data support an important role for the HERV-K *env* protein in differentiating normal and malignant melanocytes but suggest that the process is more complicated than direct up regulation of the HERV-K promoter in malignancy.

8.4 pERK signalling pathway and HERV-K *gag* protein expression

In addition to HERV-K protein expression in serum starvation, the pERK signalling pathway was investigated to reveal any correlation between HERV-K expression and the pERK pathway. The phenotype of melanoma cells was modified by treatment with a pERK inhibitor – the cells appeared to show more cell body extensions compared with control, especially at a concentration of 20 um/ml. This could be related with changes to the signalling pathways that are important for maintaining their cellular morphology. Our data also showed that HERV-K *gag* expression is likely dependent upon the pERK signalling pathway, because the expression of *gag* was significantly decreased in

melanoma cells, exposed to 20um/ml of pERK inhibitor. A recent study by (Li *et al.*, 2010) demonstrated that HERV-K *gag* and *env* protein levels were more closely associated with activation of the pERK signalling pathway in melanoma cells compared to the cells of benign nevi, by IHC.

Further research should be undertaken to investigate the level of *env* protein expression. This would be an important next step because we did not test *env* expression when pERK pathway was inhibited, due to lack of time.

8.5 Determination of methylation levels in melanoma cells

The methylation of the HERV-K 5′ LTR was detected at a low level in all melanoma cells, according to bisulphite sequence data, when grown under 10% FBS. In 1%FBS culture we observed an increase in DNA methylation at the HERV-K 5′ LTR in all melanoma cells level compared with 10%FBS culture. However, this increase in methylation level was not statistically significant. From these data, it could be suggested that HERV-K expression is influenced by two mechanisms; the methylation status of the HERV-K 5′ LTR promoter and also transcription factors that were able to initiate transcription even if their binding site is methylated. This indicates that the epigenetic modification of the HERV-K 5′ LTR does not reliably indicate the activity of the promoter, and direct measurement of mRNA and protein levels is required to dissect these effects.

8.6 GAPS and NGS characterisation of HERV-K loci in cancer cell lines.

In this thesis, we reported how the Genome-wide Amplification of Proviral Sequences (GAPS) approach, coupled with Ion Torrent NGS, can be used to simultaneously detect

proviral loci and their polymorphism status in cell lines of diverse origin. This approach was developed from the original 3′ LTR GAPS protocol and is specific for proviruses retaining the *env* gene (Macfarlane & Badge, 2015). In the results presented here GAPS and Ion Torrent NGS was applied to seven cancer cell lines of different origin (melanoma, breast, prostate, colon cancer and teratocarcinoma).

As these cell lines were all independently derived from different individuals they constitute a small diverse population sample. While we did not detect any novel HERV-K HML2 proviral insertion loci within this sample, the approach demonstrably captured a substantial proportion of known fixed and polymorphic HERV-K HML-2 proviruses, suggesting that it is suitable for population screening for novel proviruses, and is a cost-effective alternative to whole genome sequencing. An important observation was that many loci annotated as solo LTRs in the reference human genome assembly (GRCh38) were amplified in our libraries despite the specificity of the procedure for proviruses. If this is not an artefact, this suggests that a major source of variation in human genome HERV content is the retention of intact proviruses in gDNA, which are erroneously annotated as solo LTRs in reference sequences.

98 candidates detected as novel proviral loci, where these loci were annotated as solo LTRs in the current human genome reference sequence. Although not validated, this number is significant compared to the numbers of insertional polymorphic proviruses, suggesting that the number of HERV proviruses is underrepresented in the reference sequence. It may be the case that these provirus insertions are fixed in human populations, but their detection suggests that the level provirus to solo-LTR polymorphism is also much greater than the reference sequence would suggest.

Interestingly over half of the putative proviral loci reported (64 loci) were recorded in the NTera2D1 teratocarcinoma cell line, while only a few novel proviral loci were amplified from T47D. It therefore may be the case that NTera2D1 cells have the ability to produce VLPs more abundantly than other cell lines because they have a much broader cohort of proviruses, which together have retained sufficient coding capacity to generate more VLPs.

This raises the possibility that, as well as epigenetic modulation of HERV expression, some variation in HERV protein expression between cell lines results from variation in the proportion of HERV-K HML2 loci that are proviruses versus solo LTRs. Our data are in agreement with those obtained by Bieda *et al.* (2001), who found that the morphologically retrovirus-like particles (RVLP) are observed in different human teratocarcinoma cell lines.

Taking together, these results gives us a new window about the distribution pattern of HERV-K proviruses in human population. If the cell lines analyses here are representative of the human population, this suggest that some individuals retain many HERV-K HML2 loci as proviruses while the others harbour the same loci as solo LTRs. This variation has the potential to impact variation in HERV protein expression and perhaps in situations where HERV genes influence may disease progression, such as melanoma, human health.

8.7 Future work

8.7.1 Chapter 3 (*pol* expression in melanoma and other cancer cells)

It is possible to hypothesize that the *env* up regulation in SKMel5 is required to form suspension cells under starvation conditions. Therefore, further work needs to be done to establish whether knock down of *env* expression in SKMel5 adherent and suspension cells is related with the modification of phenotype or not. This could be achieved using the siRNA approach adopted by Serafino e*t al.*, (2009), in cultures of SKMel5 grown in different serum concentrations.

In order to detect whether the mRNA level of *gag* and *env* coincide with protein levels, the qPCR experiments should be repeated in melanoma and other cancer cells using specific primers for *gag* and *env* under serum starvation.

8.7.2 Chapter 4 (HERV-K *gag* and *env* protein expression and pERK signaling pathway in melanoma cells)

In regards to the pERK results, *gag* protein expression was down regulated by pERK inhibitor. Further research could also be conducted to determine the effectiveness of pERK inhibition on *env* (mRNA and protein) expression in melanoma and other cancer cells, especially breast cancer under starvation conditions. The purpose of this would be to determine whether inhibition of pERK-mediated down regulation of *env*/*gag* is specific for melanoma cells.

8.7.3 Chapter 5 (Methylation levels in HERV-K5′ LTR)

The results of methylation state analyses suggest that DNA methylation at HERV-K 5′ LTR did not show a strong effect on LTR promotor activation. Thus, further research should be undertaken to explore whether the chromatin modifications at HERV-K 5′ LTR are more important than DNA methylation. This could be carried out using Chromatin Immuno-Precipitation (ChIP) and HERV-K 5′ LTR specific PCR, to assess levels of repressive chromatin marks (for example H3K4me3) at a subset of HERV loci, or by ChIP-seq to enable genome-wide assessment of changes / difference in chromatin marks between cultures of the same cell in different conditions, or expressing different phenotypes.

8.7.4 Chapter 6 (Characterisation of profile HERV provirus by GAPS –NGS approach)

This study has also shown that the GAPS –NGS is a good approach to discover new proviral loci in gDNA. The 98 novel proviral loci discovered in this project and the profile of HERV-K provirus polymorphism may associate with variation in levels of expression. Further studies need to be carried out in order to validate, by locus specific PCR, each candidate locus. Moreover, GAPS could cover a larger area of human genome DNA by using another restriction enzyme that cuts human genomic DNA more frequently than *Ase*I, used in this study. Importantly, as only a proportion of HERV-K proviral loci are accessible to GAPS, our observation of greater prevalence of intact proviral loci than represented in the reference genome sequence is logically an under estimate of the true prevalence.

8.7.5 RNA seq analysis of differentially cultured melanocytes and melanoma cell lines.

As discussed in Chapter 7, the analysis of these experiments was constrained by time. All experimental work was completed on the planned time frame, but sufficient time was not allowed for sequencing and analysis.

The analysis completed does suggest that cultured cells do respond to serum starvation through transcriptional changes and that more detailed analysis could reveal the pathways and processes that drive these changes. The statistically defensible differences in gene expression are biologically consistent, but do not reveal specific insight into detail of the process. There are additional data that could be generated from un-sequenced libraries, and the development of the analysis pipeline will enable their integration.

Ultimately the generation of these RNA seq libraries with total RNA should allow the analysis of transcripts from non-protein coding gene, aberrant transcripts, and transposons. One approach would be to map the existing data to characterised HERV-K HML2 proviruses to measure their gene expression and compare these data to the qPCR analyses described in this thesis. The ability to quantitate HERV transcripts across the proviral genome may allow us to understand the incongruent results between protein and RNA levels of HERV genes (particularly *gag* and *env*) that we have observed in our experiments, in a less biased manner. It even may be possible to identify transcripts originating from individual HERV copies to establish whether the variation we have observed by iontorrent NGS sequencing in proviral / solo LTR occupancy is reflected at the transcriptional level, with the existing data.

Appendix1: **Map of HERVK107**

GRC37 / hg19, chr5:156087395-156092928

AACTGTTCAAAAGGGAGAAAATTTACGGATTTCTTCCACAATGGTATCAG
TCATACTGCCCCACATACCAAGTACCTACAGCTAAGAAGCTCAATATACA
CTCCTCCTTGTGTCTGTCTGTCTCTCTGTCTCTTTGTATGTGTCTTTCTC
ACTTTCAATATATATGTTTATAcagaatgaggatcctttatctgaaataa
gggaccagaagtgttttgaattatggatttttccagattttgaaatacct
gcattacaggttgagcaccccaaatctgaaaatctgaaatttgaaatgct
ccaatgagtatttcctttgagctttatgtcagtcctcaaaaagttttcaa
ttttagagcattttggattttggattttggatttgggacacccaacctg
CATGGtttttttaattttaatttttatttattattagtttttgagacaaag
tcactctgtcacccaggctggggtgcagtggcatgatctcaactcactgc
tgtggggaaaagcaagagagatcaaattgttactgtgtctgtgtagaaag
aagtagacataggagactccattttgttatgtgctaagaaaaattcttct
gccttgagattctgttaatctatgaccttacccccaaccccgtgctctct
gaaacgtgtgctgtgtcaactcagggttgaatggattaagggcggtgcag
gatgtgctttgttaaacagatgcttgaaggcagcatgctccttaagagtc
atcaccactccctaatctcaagtacccagggacacaaaaactgcggaagg
ccgcagggacctctgcctaggaaagccaggtattgtccaaggtttctccc
catgtgatagtctgaaatatggcctcgtgggaagggaaagacctgaccgt
cccccagcccgacacctgtaaagggtctgtgctgaggaggattagtaaaa
gaggaaggaatgcctcttgcagttgagacaagaggaaggcatctgtctcc
tgcctgtccctgggcaatggaatgtctcggtataaaacccgattgtatgc
tccatctactgagatagggaaaaaccgccttagggctggaggtgggacct
gcgggcagcaatactgctttgtaaagcattgagatgtttatgtgtatgca
tatccaaaagcacagcacttaatcctttacattgtctatgatgccaagac
ctttgttcacgtgtttgtctgctgaccctctccctacaattgtcttgtga
ccctgacacatccccctctttgagaaacacccacagatgatcaataaata
ctaagggaactcagaggctggcgggatcctccatatgctgaacgctggtt
ccccgggtcccctttatttctttctctatactttgtctctgtgtctttttc
ttttccaaatctctcgtcccaccttacgagaaacacccacaggtgtgtag
gggcaacccaccccctacatctggtgcccaacgtggaggcttttctctagg
gtgaaggtacgctcgagcgtggtcattgaggacaagtcgacgagagatcc
cgagtacatctacagtcagccttacggtaagcttgcgcgctcggaagaag
ctagggtgataatggggcaaactaaaagtaaaattaaaagtaaatatgcc
tcttatctcagctttattaaaattcttttaaaaagaggggagttaaagt
atctacaaaaaatctaatcaagctatttcaaataatagaacaattttgcc
catggtttccagaacaaggaactttagatctaaaagattggaaaagaatt
ggtaaggaactaaaacaagcaggtaggaagggtaatatcattccacttac
agtatggaatgattgggccattattaaagcagctttagaaccatttcaaa
cagaagaagatagcatttcagtttctgatgcccctggaagctgtttaata
gattgtaatgaaaacacaaggaaaaaatcccagaaagaaaccgaaagttt
acattgcgaatatgtagcagagccggtaatggctcagtcaacgcaaaatg
ttgactataatcaattacaggaggtgatatatcctgaaacgttaaaatta
gaaggaaaaggtccaga*ATTAAT*ggggccatcagagtctaaaccacgagg
cacaagtcctcttccagcaggtcaggtgcccgtaagattacaacctcaaa
agcaggttaaagaaaataagacccaaccgccagtagcctatcaatactgg
ccgccggctgaacttcagtatcggccacccccagaaagtcagtatggata
tccaggaatgccccagcaccacagggcagggcgccataccatcagccgc
ccactaggagacttaatcctatggcaccacctagtagacagggtagtgaa
ttacatgaaattattgataaatcaagaaaggaaggagatactgaggcatg
gcaattcccagtaacgttagaaccgatgccacctggagaaggagcccaag
agggagagcctcccacagttgaggccagatacaagtctttttcgataaaa
atgctaaaagatatgaaagagggagtaaaacagtatggacccaactcccc
ttatatgaggacattattagattccattgcttatggacatagactcattc

cttatgattgggagattctggcaaaatcgtctctctcaccctctcaattt
ttacaatttaagacttggtggattgatggggtacaagaacaggtccgaag
aaatagggctgccaatcctccagttaacatagatgcagatcaactattag
gaataggtcaaaattggagtactattagtcaacaagc*ATTAAT*gcaaaat
gaggccattgagcaagttagagctatctgccttagagcctgggaaaaaat
ccaagacccaggaagtacctgcccctcatttaatacagtaagacaaggtt
caaaagagccctaccctgattttgtggcaaggctccaagatgttgctcaa
aagtcaattgccgatgaaaaagccggtaaggtcatagtggagttgatggc
atatgaaaacgccaatcctgagtgtcaatcagccattaagccattaaaag
gaaaggttcctgcaggatcagatgtaatctcagaatatgtaaaagcctgt
gatggaatcggaggagctatgcataaagctatgcttatggctcaagcaat
aacaggagttgttttaggaggacaagttagaacatttggaggaaaatgtt
ataattgtggtcaaattggtcacttaaaaaagaattgcccagtcttaaac
aaacagaatataactattcaagcaactacaacaggtagagagccacctga
cttatgtccaagatgtaaaaaaggaaaacattgggctagtcaatgtcgtt
ctaaatttgataaaaatgggcaaccattgtcgggaaacgagcaaaggggc
cagcctcaggccccacaacaaactggggcattcccaattcagccatttgt
tcctcagggttttcagggacaacaacccctactgtcccaagtgtttcagg
gaataagccagttaccacaatacaacaattgtccctcaccacaagcggca
gtgcagcagtagatttatgtactatacaagcagtctctctgcttccaggg
gagcccccacaaaaaatccctacaggggtatatggcccactgcctgaggg
gactgtaggactaatcttgggaagatcaagtctaaatctaaaaggagttc
aaattcatactagtgtggttgattcagactataaaggcgaaattcaattg
gttattagctcttcaattccttggagtgccagtccaagagacaggattgc
tcaattattactcctgccatatattaagggtggaaatagtgaaataaaaa
gaataggagggcttgtaagcactgatccaacaggaaaggctgcatattgg
gcaagtcaggtctcagagaacagacctgtgtgtaaggccattattcaagg
aaaacagtttgaagggttggtagacactggagcagatgtctctatcattg
ctttaaatcagtggccaaaaaattggcctaaacaaaaggctgttacagga
cttgtcggcataggcacagcctcagaagtgtatcaaagtatggagatttt
acattgcttagggccagataatcaagaaagtactgttcagccaatgatta
cttcaattcctcttaatctgtggggtcgagatttattacaacaatggggt
gcggaaatcaccatgcccgctccattatatagccccacgagtcaaaaaat
catgaccaagatgggatatataccaggaaagggactagggaaaaatgaag
atggcattaaagttccagttgaggctaaaataaatcaagaaagagaagga
atagggtatccttttagggggcggtcactgtagagcctcctaaacccata
ccactaacttggaaaacagaaaaaccggtgtgggtaaatcagtggccgct
accaaaacaaaaactggaggctttacatttattagcaaatgaacagttag
aaaagggtcacattgagccttcgttctcaccttggaattctcctgtgttt
gtaattcagaagaaatcaggcaaatggcgtatgttaactgacttaagggc
tgtaaacgccgtaattcaacccatggggcctctccaacccgggttgccct
ctccggccatgatcccaaaagattggcctttaattataattgatctaaag
gattgctttttaccatccctctggcagagcaggattgtgaaaaatttgc
ctttactataccagccataaataataaagaaccagccaccaggtttcagt
ggaaagtgttacctcagggaatgcttaatagtccaactatttgtcagact
tttgtaggtcgagctcttcaaccagtgagagaaaagttttcagactgtta
tattattcattatattgatgatattttatgtgctgcagaaacgaaagata
a*ATTAAT*tgactgttatacatttctgcaagcagaggttgccaatgctgga
ctggcaatagcatctgataagatccaaacctctactccttttcattattt
agggatgcagatagaaaatagaaaaattaagccacaaaaaatagaaataa
gaaaagacacattaaaaacactaaatgattttcaaaaattactaggagat
*ATTAAT*tggattcggccaactctaggcattcctacttatgccatgtcaaa
tttgttctctatcttaagaggagactcagacttaaatagtcaaagaatat
taaccccagaggcaacaaaagaaattaaattagtggaagaaaaaattcag
tcagcgcaaataaatagaatagatcccttagccccactccaacttttat
ttttgccactgcacattctccaacaggcatcattattcaaaatactgatc
ttgtggagtggtcattccttcctcacagtacagttaagacttttacattg
tacttggatcaaatagctac*ATTAAT*cggtcagacaagattacgaataac

aaaattatgtggaaatgacccagacaaaatagttgtccctttaaccaagg
aacaagttagacaagcctttatcaattctggtgcatggcagattggtctt
gctaattttgtgggacttattgataatcattacccaaaaacaaagatctt
ccagttcttaaaattgactacttggattctacctaaaattaccagacgtg
aacctttagaaaatgctctaacagtatttactgatggttccagcaatgga
aaagcagcttacacagggccgaaagaacgagtaatcaaaactccatatca
atcggctcaaagagcagagttggttgcagtcattacagtgttacaagatt
ttgaccaacctatcaatattatatcagattctgcatatgtagtacaggct
acaagggatgttgagacagctctaattaaatatagcatggatgatcagtt
aaaccagctattcaatttattacaacaaactgtaagaaaaagaaatttcc
cattttatattactcatattcgagcacacactaatttaccagggcctttg
actaaagcaaatgaacaagctgacttactggtatcatctgcactcataaa
agcacaagaacttcatgctttgactcatgtaaatgcagcaggattaaaaa
acaaatttgatgtcacatggaaacaggcaaaagatattgtacaacattgc
acccagtgtcaagtcttacacctgcccactcaagaggcaggagttaatcc
cagaggtctgtgtcctaatgcattatggcaaatggatgtcacgcatgtac
cttcatttggaagattatcatatgttcatgtaacagttgatacttattca
catttcatatgggcaacttgccaaacaggagaaagtacttcccatgttaa
aaaacatttattgtcttgttttgctgtaatgggagttccagaaaaaatca
aaactgacaatggaccaggatattgtagtaaagctttccaaaaattctta
agtcagtggaaaatttcacatacaacaggaattccttataattcccaagg
acaggccatagttgaaagaactaatagaacactcaaaactcaattagtta
aacaaaaagaaggggagacagtaaggagtgtaccactcctcagatgcaa
cttaatctagcactctatactttaaatttttaaacatttatagaaatca
gactactacttctgcagaacaacatcttactggtaaaaagaacagcccac
atgaaggaaaactaatttggtggaaagataataaaaataagacatgggaa
atagggaaggtgataacgtgggggagaggttttgcttgtgtttcaccagg
agaaaatcagcttcctgtttggttacccactagacatttgaagttctaca
atgaacccatcggagatgcaaagaaaagggcctccacggagatggtaaca
ccagtcacatggatggataatcctatagaagtatatgttaatgatagtgt
atgggtacctggccccatagatgatcgctgccctgccaaacctgaggaag
aagggatgatgataaatatttccattgggtatcgttatcctcctatttgc
ctagggagagcaccaggatgtttaatgcctgcagtccaaaattggttggt
agaagtacctactgtcagtcccatcagtagattcacttatcacatggtaa
gcgggatgtcactcaggccacgggtaaattatttacaagacttttcttat
caaagatcattaaaatttagacctaaagggaaaccttgccccaaggaaat
tcccaaagaatcaaaaaatacagaagttttagtttgggaagaatgtgtgg
ccaatagtgcggtgatattataaaacaatgaatttggaactattatagat
tgggcacctcgaggtcaattctaccacaattgctcaggacaaactcagtc
gtgtccaagtgcacaagtgagtccagctgttgatagcgacttaacagaaa
gtttagacaaacataagcataaaaaattgcagtctttctacccttgggaa
tggggagaaaaaggaatctctaccccaagaccaaaaatagtaagtcctgt
ttctggtcctgaacatccagaattatggaggcttactgtggcctcacacc
acattagaatttggtctggaaatcaaactttagaaacaagagattgtaag
ccattttatactgtcgacctaaattccagtctaacagttcctttacaaag
ttgcgtaaagccccttatatgctagttgtaggaaatatagttattaaac
cagactcccagactataacctgtgaaaattgtagattgcttacttgcatt
gattcaacttttaattggcaacaccgtattctgctggtgagagcaagaga
gggcgtgtggatccctgtgtccatggaccgaccgtgggaggcctcaccat
ccgtccatattttgactgaagtattaaaaggtgttttaaatagatccaaa
agattcatttttacttttaattgcagtgattatggg*ATTAAT*tgcagtcac
agctacggctgctgtagcaggagttgcattgcactcttctgttcagtcag
taaactttgttaatgattggcaaaagaattctacaagattgtggaattca
caatctagtattgatcaaaaattggcaaatcaa*ATTAAT*gatcttagaca
aactgtcatttggatgggagacagactcatgagcttagaacatcgtttcc
agttacaatgtgactggaatacgtcagatttttgtattacaccccaaatt
tataatgagtctgagcatcactgggacatggttagacgccatctacaggg
aagagaagataatctcactttagacatttccaaattaaaagaacaaattt

tcgaagcatcaaaagcccatttaaatttggtgccaggaactgaggcaatt
gcaggagttgctgatggcctcgcaaatcttaaccctgtcacttgggttaa
gaccattggaagtactacgattataaatctcat*ATTAAT*ccttgtgtgcc
tgttttgtctgttgttagtctgcaggtgtacccaacagctccgaagagac
agcgaccatcgagaacgggccatgatgatgatggcggttttgtcgaaaag
aaaaggggggaaatgtggggaaaagcaagagagatcaaattgttactgtgt
ctgtgtagaaagaagtagacataggagactccattttgttatgtgctaag
aaaaattcttctgccttgagattctgttaatctatgaccttaccccaaac
cccgtgctctctgaaacgtgtgctgtgtcaactcagggttgaatggatta
agggcggtgcaggatgtgctttgttaaacagatgcttgaaggcagcatgc
tccttaagagtcatcaccactccctaatctcaagtacccagggacacaaa
aactgcggaaggccgcagggacctctgcctaggaaagccaggtattgtcc
aaggtttctccccatgtgatagtctgaaatatggcctcgtgggaagggaa
agacctgaccgtcccccagcccgacacctgtaaagggtctgtgctgagga
ggattagtaaaagaggaaggaatgcctcttgcagttgagacaagaggaag
gcatctgtctcctgcctgtccctgggcaatggaatgtctcggtataaaac
ccgattgtatgctccatctactgagatagggaaaaaccgccttagggctg
gaggtgggacctgcgggcagcaatactgctttgtaaagcattgagatgtt
tatgtgtatgcatatccaaaagcacagcacttaatcctttacattgtcta
tgatgccaagacctttgttcacgtgtttgtctgctgaccctctccccaca
attgtcttgtgaccctgacacatccccctctttgagaaacacccacagat
gatcaataaatactaagggaactcagaggctggcgggatcctccatatgc
tgaacgctggttccccgggtcccctttatttctttctctatactttgtctc
tgtgtctttttctttttccaaatctctcgtcccaccttacgagaaacaccc
acaggtgtgtaggggcaacccacccctaca*actgc*aacctccgcctccta
gttcaagtgattctcctgcctcagcctcccaactagctgggattaccggc
atgcaccaccacacccagctaatttttgtatttttagtagagatggggtt
tcaccatggtggccaggctgatctggaactcctgagctcaggtgatcctc
ctgcctcagcctcccaaagtgctgggatcataggcatgagccaccccaca
cctggccTCAACCTGTATGTTTTTaatgagaacacatggacacaggaagg
ggaacatcacactccggggactactgtggggtgggggggagggggggagga
tagcattaggagatatacctaatgctaaatgaccagttaatgggtgcagc
acaccaacatggcacatgtatacatatgtaacaaacctgcacattgtgca
catgtaccctaaaacttaatgtataataataataaaaTAAAATAAAAGAG
AGAGGAAAAAAAAAAAGTCAGACTATTTATCCAATATTTTGGTACTGTAA
CTGTCAGTGGTGTGCGGAGAATTACCAAGCCTCTA*ATTAAT*TCACCCCTA
AGGCTCTAATAGACTGTTTTGATA

Appendix 2:

Table shows the data of colony formation assay (plate efficiency) to melanoma and other cancer cell lines with three biological repeats.

Name of cell lines	Plate efficiency (%) for each biological repeats in 10% FBS			Plate efficiency (%) for each biological repeats in 1% FBS		
	B1	**B2**	**B3**	**B1**	**B2**	**B3**
A375P	54.4	51.7	50.9	0	0	0
A375M	35.9	32.5	32.7	0	0	0
SKMel5	25	22.1	30.7	7.1	7.4	9.7
SKMel28	15.3	14.7	14.8	7	5.3	7
LNCAP	18.5	18	18.5	4.5	5	4.8
T47D	56.2	52.1	40.8	12.3	13.5	12.5
SW480	47.8	47.1	46.6	13.5	11.3	11.3

Appendix 3

The table shows 32 loci (from hg38 coordinates and intersect data) were detected overlapped with previously published (Subramanian *et al*., 2011; Wildschutte *et al*., 2016).

Chromosome	Start	End	Locus
chr1	75377910	75378335	1p31.1
chr1	160699785	160700602	1q23.3
chr1	166604957	166605521	1q24.1
chr10	6823149	6824384	10p14
chr10	26894225	26895280	10p12.1
chr10	99819570	99821038	10q24.2
chr11	74127714	74128313	11q13.4
chr11	62858031	62858748	11q12.3
chr11	57432668	57433333	11q12.1
chr11	41305883	41306186	11p12
chr11	10391107	10391457	11p15.4
chr12	7900824	7901291	12p13.31
chr12	58325898	58327655	12q14.1
chr19	387494	388712	19p13.3
chr2	129961543	129962251	2q21.1
chr3	101700853	101701564	3q12.3
chr3	113023287	113024452	3q13.2
chr3	185561475	185562714	3q27.2
chr4	151000	151912	4p16.3
chr4	156304499	156304990	4q32.1
chr5	30486412	30486899	5p13.3
chr5	43581132	43582071	5p12
chr5	156656294	156657938	5q33.3
chr6	44327527	44328189	6p21.31
chr7	4581380	4582636	7p22.1
chr7	140552819	140553143	7q34
chr8	139458248	139460111	8q24.3
chr1	160699871	160700454	1q23.3
chr10	26894293	26894635	10p12.1
chr2	129961608	129962167	2q21.1
chr5	30486482	30486848	5p13.3
chr5	43664880	43665125	5p12

Appendix4

This table shows 16 loci (from hg38 coordinates and intersect data) were detected as novel loci.

Chromosome	Start	End	Locus
chr2	37226141	37226856	2p22.2
chr2	55280093	55280833	2p16.1
chr5	18579281	18580123	5p14.3
chr6	111254751	111255196	6q21
chr6	137845272	137846135	6q23.3
chr6	93174147	93174426	6q16.1
chr7	27741087	27741833	7p15.2
chr7	123781458	123781867	7q31.32
chr7	125220548	125221313	7q31.33
chr11	41305883	41306186	11p12
chr12	7900824	7901291	12p13.31
chr13	42967290	42968062	13q14.11
chr16	74799147	74799564	16q23.1
chr22	11550440	11550904	22p11.2
chrX	123680426	123681093	xq25
chrY	19699196	19699623	Yq11.223

Appendix 5

This table shows 64/98 loci (from hg38 coordinates and subtract data) were detected in specific cell as novel loci.

chromosome	start	end	specific cell
chr12	29931490	29932199	Ntera 2D1
chr12	29931450	29932245	Ntera 2D1
chr12	8462756	8463347	Ntera 2D1
chr7	27741201	27741796	Ntera 2D1
chr7	125220522	125221314	Ntera 2D1
chr6	52923468	52923947	Ntera 2D1
chr7	125220622	125221320	Ntera 2D1
chr1	46330295	46330885	Ntera 2D1
chr6	44327563	44328247	Ntera 2D1
chr1	46330372	46330856	Ntera 2D1
chr2	55280098	55280828	Ntera 2D1
chr6	111254733	111255186	Ntera 2D1
chr3	101703263	101703893	Ntera 2D1
chr1	29337735	29338249	Ntera 2D1
chr16	74799211	74799538	Ntera 2D1
chr2	192541525	192542084	Ntera 2D1
chr11	24446995	24447310	Ntera 2D1
chr4	72128680	72129279	Ntera 2D1
chr12	9601363	9602330	Ntera 2D1
chr2	37751431	37752004	LNCAP
chr9	95269468	95269913	LNCAP
chr12	56400250	56400522	LNCAP
chr19	28131489	28132367	LNCAP
chr16	47865285	47866757	LNCAP
chr2	194503812	194504460	lNCAP
chr5	93455659	93456718	LNCAP
chr6	111254751	111255196	LNCAP
chr2	37751350	37752072	LNCAP
chr18	2000727	2000982	LNCAP
chr6	24662653	24663157	LNCAP
chr6	93174187	93174424	LNCAP
chr11	41305883	41306186	LNCAP
chr2	37226093	37226925	A375
chr11	62858031	62858748	A375
chr5	18579254	18580160	A375

chr4	151046	151870	A375
chr1	29337752	29338278	A375
chr4	151066	151886	A375
chr7	27741087	27741833	A375
chr4	151088	151924	A375
chr15	101859288	101860155	A375
chr7	125220605	125221312	A375
chr6	44327606	44328201	A375
chr17	54202425	54203462	A375
chr6	93174147	93174426	A375
chr4	119342090	119342679	SW480
chr6	33809287	33810151	SW480
chr1	46330384	46330856	SW480
chr7	125220497	125221356	SW480
chr13	76191565	76192191	SW480
chr11	72165198	72166253	SW480
chr12	56400149	56400544	SW480
chr2	37226059	37226923	SKMEL28
chr22	11550440	11550904	SKMEL28
chr12	51453646	51454491	SKMEL28
chr11	62858012	62858754	SKMEL28
chr3	175905503	175905745	SKMEL28
chr16	74799147	74799564	SKMEL5
chr6	125775685	125776005	SKMEL5
chr7	123781458	123781867	SKMEL5
chr7	123781463	123781814	SKMEL5
chr7	80266520	80267026	skMEL5
chr14	20085396	20086041	T47D
chr10	65408795	65409370	T47D

This table shows 32/98 loci (from hg38 coordinates and subtract data) were shared with previously published (Subramanian *et al*., 2011; Wildschutte *et al*., 2016).

Chromosome	Start	End
chr1	66424125	66424822
chr1	29337793	29338221
chr2	37226141	37226856
chr2	55280093	55280833
chr2	37226138	37226808
chr2	37226144	37226800
chr2	37226144	37226772
chr2	37226142	37226816
chr5	43581247	43582060
chr5	18579281	18580123
chr5	43581313	43582103
chr5	18579279	18580054
chr5	43581157	43582126
chr5	43581132	43582071
chr5	18579353	18580060
chr6	137845272	137846135
chr6	137845298	137846115
chr6	159794679	159794880
chr6	44327612	44328217
chr7	27741173	27741868
chr7	125220548	125221313
chr7	125220527	125221348
chr7	125220549	125221379
chr11	10391107	10391457
chr12	8462773	8463346
chr12	29931490	29932220
chr12	9601262	9602258
chr7	125220590	125221313
chrX	125780379	125781081
chrX	123680426	123681093
chrX	123680467	123681089
chrX	123680488	123681185

This table shows2/98 loci (from hg38 coordinates and subtract data) were shared with in two cell lines.

Chromosome	Start	End	Cells
chrY	19699192	19699656	LNCAP and Ntera 2D1
chr4	119342187	119342640	LNCAP and SW480

Appendix 6

This table shows 9 loci (from hg38 coordinates and subtract data) were shared with previously published (Subramanian *et al*., 2011; Wildschutte *et al*., 2016).

Chromosome	**Start**	**End**	**Locus**	**Name of paper**	**Name of cell**
chr5	43581217	43582210	5p12_subt1	Subramanian	A375P
chr7	140552819	140553143	7q34_subt1	Subramanian	LNCAPL
chr4	156304499	156304990	4q32.1_subt1	Subramanian	SKMEL28
chr6	44327606	44328188	6p21.1_subt1	Subramanian	Ntera 2D1
chr11	62858073	62858661	q12.3	Subramanian	A375ML
chr4	156304659	156304905	4q32.1_subt1	Subramanian	SKMEL28
chr14	20085396	20086041	14q11.2_subt2	Subramanian	T47D
chr3	101703125	101704073	3q12.3_subt1	Subramanian	ALL
chr1	219389867	219390564	q41	Wildschutte	LNCAP

.

Appendix 7

Table shows differential expressed (up and down) genes comparing SKMEL28 in 1% FBS versus SKMEL28 in 10% FBS.

Gene Symbol	p-value	Fold Change
SPINK1	6.00E-19	74.1191
C5orf47	1.60E-06	15.5082
LINC00324	2.06E-17	10.0312
PCDH20	9.06E-15	4.98867
RNF152	5.03E-05	4.56713
PDE4C	0.000148573	4.46808
HMGCS1	2.19E-18	4.42576
VWDE	4.35E-09	4.35037
SCGB1D2	1.49E-06	4.20855
FCRLA	6.63E-26	3.54914
TFRC	1.76E-15	3.46223
MSMO1	1.11E-23	3.42048
LOC107984069	7.64E-09	3.29025
TMEM97	2.79E-24	3.26573
TM7SF2	7.00E-07	3.1887
UBBP4	0.000434209	3.11886
INSIG1	1.54E-06	3.09777
GCNT3	1.04E-07	3.06061
MVD	5.54E-16	3.00706
LOC101060207	1.17E-09	2.95477
HMGCR	5.13E-14	2.89161
IDI1	5.96E-13	2.87103
MDGA1	1.09E-17	2.86314
SNORD65	0.000194808	2.817
FDPS	7.13E-17	2.63177
LINC00890	3.98E-18	2.57842
DHCR7	1.96E-07	2.55276
STARD4	7.79E-12	2.54397
ATP10D	2.84E-14	2.513
CYP51A1P2	1.02E-06	2.44317
IQCE	0.000647246	2.43165
LSS	1.37E-09	2.42937
SQLE	1.77E-14	2.41037
RASD1	1.60E-06	2.3551
TRS-ACT1-1	0.000198933	2.35017
ACAT2	1.54E-15	2.29603
NMRK2	7.85E-06	2.27169
NAMA	0.000105353	2.25973
SNORA37	0.000176832	2.21962
STARD10	1.46E-05	2.19141
SNORD12	4.84E-06	2.17711
C11orf96	9.81E-08	2.14365
AMD1	7.00E-05	2.13277
LINC01465	7.36E-05	2.07084
SNORA12	3.47E-06	-2.02115
HIST1H4K	5.74E-05	-2.12328

ZNF674	8.44E-08	-2.16845
TRNY	0.000371416	-2.20336
RNU5E-1	1.39E-09	-2.23891
EXPH5	3.89E-07	-2.25063
SEPN1	4.60E-11	-2.31358
RNU1-1	6.27E-06	-2.32366
RNU1-3	6.27E-06	-2.32366
RNVU1-18	6.27E-06	-2.32366
RNU1-28P	5.36E-06	-2.33376
RNU1-2	8.67E-06	-2.34172
RNU1-4	8.67E-06	-2.34172
RNU1-27P	7.42E-06	-2.35219
TRNA	8.01E-05	-2.51219
COL4A4	9.80E-11	-2.56359
TNFRSF12A	1.98E-10	-2.56621
WDR74	3.81E-13	-2.59399
RNU11	3.94E-10	-2.61388
TRNN	2.32E-05	-2.63227
MTCO1P2	6.77E-10	-2.82723
RPPH1	4.11E-12	-3.00332
SCARNA1	1.01E-08	-3.05134
TNFRSF13C	0.000212037	-3.0868
WARS	5.67E-05	-3.13367
LOC441907	3.56E-05	-3.13753
RNU2-2P	4.30E-16	-3.1693
MIR100HG	9.43E-06	-3.31509
DDIT4	9.75E-16	-3.3371
SNORA75	0.000135184	-3.50231
SNORA23	2.26E-11	-3.51133
LOC107984008	0.000632239	-3.51748
PER2	5.43E-08	-3.52236
SNORA74A	3.92E-13	-3.70233
RNU4-1	0.00023649	-3.89727
SCARNA10	1.30E-16	-3.93306
RNU12	3.17E-08	-3.94557
GEM	2.56E-19	-3.99113
TRNC	1.88E-10	-4.14151
RNU5D-1	2.27E-07	-4.85141
PCK2	2.18E-06	-4.95792
SCARNA5	4.03E-25	-5.05402
SNORA74B	2.41E-14	-5.72242
FAM129A	4.94E-05	-5.93883
SNORA81	1.75E-19	-6.1197
DUX4L19	7.11E-06	-6.47973
ATF3	0.000108975	-9.72397
SNORA53	9.67E-15	-10.0848

Table shows top 100 upregulated genes comparing Melanocytes and SKMEL28 in 10%FBS.

Gene Symbol	p-value	Fold Change
LOC107986133	0.00148306	67017000000
LOC102723438	1.18356E-05	17177200000
LOC102724042	1.18356E-05	17177200000
CES1	1.74419E-19	10688100000
FAM27E2	1.81945E-13	2992280000
KRT18P48	0.000020457	857449000
GPAA1P2	0.00223233	602004000
LOC105375355	0.000761268	118807000
LOC101928718	0.00262624	19960300
LOC101926969	2.13123E-18	5407830
LOC107987427	1.11645E-13	757756
NBPF4	2.09996E-23	52088.3
CCDC169-SOHLH2	0.00101975	39005.3
XIST	6.32853E-93	34136.3
PHBP6	9.76236E-06	14199.1
LOC105379447	0.00078009	6960.7
CADM3	2.20622E-80	6641.2
MAL	3.44618E-73	6199.32
EPS8L2	2.85669E-78	4070.27
COL1A2	1.3718E-99	3380.55
CYP1B1	1.62208E-78	2993.3
LOC105373249	8.4183E-06	2837.4
PXDN	4.675E-99	2835.95
IGFBP4	1.22873E-72	2572.27
WFDC1	8.71675E-70	2037.37
KCNE4	7.52835E-87	1831.94
LOC105374875	2.65993E-80	1763.2
SECTM1	8.96529E-67	1734.37
LOC107986350	1.00694E-37	1693.56
LOC392334	6.59426E-06	1674.37
TGFBI	4.16241E-73	1631.11
TSPYL5	4.2547E-87	1578.08
AEBP1	3.82874E-71	1535.87
FCGRT	2.70046E-77	1517.89
IRAK3	1.06315E-83	1512.04
CDR1	1.76894E-61	1466.48
CNN2P2	0.000153544	1299.38
COL4A1	3.96705E-40	1136.55
TMEM132D	8.87449E-97	1072.52
RAMP1	5.99661E-67	966.528
PTCHD4	4.99459E-72	896.61
CALN1	5.4678E-65	860.977
KCNA2	2.80798E-82	819.535
IFI27	5.71481E-55	809.99
COL4A2	8.52534E-65	748.988
DLGAP2	4.10344E-68	686.071
FES	1.27526E-77	677.782
CSRNP3	4.11845E-81	668.625
MID2	3.69637E-71	663.705

ELFN1	1.42055E-76	600.648
PCDH11X	2.58975E-75	588.305
KIT	9.83174E-78	571.647
LRRC2	2.09527E-74	524.591
LOC105372698	1.35992E-68	510.827
GRIA3	4.86637E-61	502.048
SOX2	2.35656E-65	485.345
PLPP7	5.08898E-54	468.236
LGI3	5.0076E-87	465.238
NPM2	2.99808E-63	452.199
LOC107986813	9.94363E-38	447.523
ZNF536	1.33034E-15	427.033
ZNF667	3.41924E-64	411.316
SALL4	1.2949E-71	404.322
LINC00654	1.97147E-77	398.998
LOC105373591	8.87977E-52	393.364
ARHGAP23	5.66599E-54	392.376
MALL	2.7881E-42	389.555
CCL18	3.65936E-61	388.52
C15orf59	2.39642E-69	383.416
HSPB2	4.11315E-57	381.454
MCF2L	1.65938E-12	379.405
ARHGEF6	3.35109E-80	371.152
IFI44L	4.61665E-63	352.048
ZNF471	9.94152E-69	350.381
OLFML2B	1.50801E-46	349.697
PNMAL1	2.4514E-45	349.442
EFNB2	1.7276E-76	349.077
SERTAD4	4.5384E-55	348.758
NYNRIN	1.66002E-68	344.736
SH3RF3	2.39922E-44	342.796
LOC105372210	1.32827E-12	332.632
BCAN	1.95113E-72	328.238
NPFFR2	8.15379E-40	327.73
SDK1	1.0305E-68	325.593
COL18A1-AS1	2.31386E-44	321.276
PNMAL2	1.94429E-59	320.477
PDE7B	4.35206E-09	309.34
KIAA1755	1.47095E-54	308.864
CRYAB	1.24004E-44	307.441
ITM2A	0.00312345	303.781
TMEM47	2.56523E-67	301.536
ZSCAN18	1.07474E-29	295.587
ZNF606	2.85356E-53	292.467
IGFBP5	1.00856E-65	285.464
SYNPO	3.96382E-79	279.131
CYGB	2.32244E-62	273.82
CDC42EP1	6.69122E-10	262.459
PNP	0.000011411	254.202
DOCK8	5.5696E-69	250.616
SLCO1C1	3.54866E-28	248.86

Table shows bottom 100 downregulated genes comparing Melanocytes and SKMEL28 in 10%FBS.

Gene Symbol	p-value	Fold Change
GUCA1C	8.09697E-28	-26.3956
LOC285303	1.8403E-07	-26.3973
LOC105376188	3.30754E-08	-26.4548
GSTT1-AS1	2.8779E-08	-26.5881
LOC105375014	1.50804E-08	-26.7918
LOC105371824	9.15632E-07	-27.0646
LOC728853	2.68015E-13	-27.1758
LOC105377441	0.00168362	-27.1827
PRDM8	1.1701E-19	-27.2161
LOC105369828	0.00063959	-27.2861
LOC100127917	1.82636E-06	-27.5684
MICALCL	0.00042511	-27.7285
CEACAM21	6.66652E-07	-28.1416
SNAR-E	0.00255208	-28.4684
TRP-TGG3-2	0.00276354	-28.6014
SERPINB2	8.29512E-16	-28.6833
APOC4	4.59324E-05	-28.8358
POU5F1B	9.79558E-07	-29.3988
ITLN2	4.05453E-07	-29.4686
TFCP2L1	2.50548E-18	-29.6648
LIN28AP1	0.000204437	-29.8114
RN7SKP42	1.90809E-07	-30.0766
PAX8	9.17096E-07	-30.2369
SCARNA3	4.59721E-05	-30.6208
ASIC5	6.55908E-08	-30.8143
UNGP2	9.92716E-15	-30.9541
NOVA1-AS1	2.22602E-09	-31.1269
LOC645553	1.2929E-09	-31.6482
LINC00328-2P	9.37912E-05	-32.0412
LOC107984616	3.94223E-05	-32.197
HP08777	5.59511E-19	-32.6296
RN7SKP194	3.00005E-20	-33.0173
LOC105373645	7.31908E-18	-33.1808
LOC105372325	1.12095E-05	-33.4018
ZNF716	6.64205E-05	-34.051
LOC105378776	0.00192709	-34.5597
LOC105378260	3.71935E-22	-34.5875
CALR3	0.000662689	-34.6916
LOC102723471	3.67781E-12	-34.7015
LOC105373848	0.000268871	-35.3623
CYP4F12	2.45177E-23	-35.7789
MIR1-1HG	9.02831E-05	-35.844
CCDC13-AS1	0.000127096	-36.4031
FRMPD2B	4.30387E-09	-36.6702
LOC105373617	5.63166E-12	-36.8475
PKP2	0.000179778	-37.1682
TRL-TAG3-1	0.000876591	-37.8344
DSG2-AS1	1.00003E-06	-38.1765
SNAR-F	1.02259E-10	-38.3557
LOC107986066	3.74899E-12	-39.1936

RN7SL736P	1.61719E-06	-39.3159
LOC105373643	2.18242E-08	-39.5011
S1PR5	0.000135786	-40.1347
OR5J2	0.000244644	-40.1976
SCARNA11	0.000394896	-41.4596
COX7BP1	4.59038E-05	-41.6305
LOC730668	7.53622E-05	-41.8138
RPL23AP88	2.09021E-09	-41.9325
LOC105369806	2.81335E-27	-42.4117
TRG-CCC1-2	0.00236202	-43.2886
CEACAM3	9.36082E-11	-43.5921
ANO2	2.2887E-06	-44.159
LOC105377712	1.69082E-06	-44.2802
RNU1-11P	2.20432E-15	-44.8995
APOA4	3.65879E-13	-45.6372
LOC105370969	0.000338196	-46.006
LOC101927865	0.000275976	-47.44
LOC107985931	1.96372E-08	-48.393
LOC151484	1.7975E-09	-51.1927
CCKBR	0.00313233	-51.4779
LOC339260	0.000140461	-51.4969
UBE2E1-AS1	0.00270731	-51.696
LOC105377367	4.38876E-08	-54.2767
EIF3FP2	2.42129E-13	-54.8022
LOC107985025	8.04185E-18	-55.4787
LOC100507661	1.04215E-14	-55.8688
IL9RP1	3.56608E-11	-56.4758
IL15RA	7.18278E-05	-57.1714
LOC105370315	0.000342014	-58.1689
RBBP4P3	1.39842E-11	-59.9358
PLCXD3	0.000822175	-63.6337
P3H2-AS1	2.72492E-13	-73.764
LOC105371529	6.51488E-07	-77.9695
CST2	5.63761E-09	-78.0007
LOC100420095	0.000560368	-80.7726
LOC107983951	2.33313E-06	-81.4441
SNORA51	3.42168E-09	-82.8993
LOC105370421	2.85998E-15	-88.1137
LOC100420250	2.05655E-05	-96.3624
KCNQ5-AS1	1.8317E-10	-104.948
LOC107987356	9.99787E-10	-1735.57
RPL7P50	0.00128152	-1837.16
ZNF670-ZNF695	0.00124063	-1876.04
LOC105371108	2.15779E-07	-27074.4
CYP2D7BP	0.000958047	-29443.4
ESPNP	4.60677E-05	-36149.5
MIR1282	0.0026139	-44061.9
LINC00431	4.04101E-07	-46115.2
TMEM110-MUSTN1	0.00314348	-278895
LRRC24	0.000446189	-228564000

Table shows top 100 upregulated genes comparing Melanocytes and SKMEL5 in 10%FBS

Gene Symbol	p-value	Fold Change
FAM27E2	2.97128E-16	18305700000
CES1	2.78275E-23	6185570000
SERHL	0.000926585	4542660000
NANOGP8	3.74822E-05	191081000
FAM13C	0.00114536	41381300
LOC442064	7.93661E-17	39699600
LIMS3-LOC440895	0.000918122	30946800
CTAGE8	4.39253E-06	4715670
TMIGD3	9.43538E-06	249657
PLA2G7	6.26568E-06	208781
CCDC169-SOHLH2	0.000209215	153352
LOC286297	6.67522E-07	56178.8
XIST	3.9219E-101	16933.3
MTAP	7.47409E-42	15698.9
CADM3	2.40198E-88	13358.3
COL11A1	2.57655E-86	12354.4
LOC101928605	1.69519E-11	11331.5
PRDX2	1.11624E-81	9710.49
TMEM132D	4.196E-105	9660.16
FXYD3	1.38216E-55	7391.85
SGIP1	1.75635E-76	6053.14
TRA2A	1.64457E-48	5632.21
PEG3	1.95476E-90	5348.2
IFI27	4.43404E-62	4298.94
PDLIM4	3.49032E-17	4181.92
LOC100499484-C9ORF174	1.37668E-55	3574.72
MX2	7.52522E-77	3369.35
CPQ	2.80367E-54	3367.06
PPP1R3C	1.06131E-80	2863.61
CDH3	2.84647E-92	2825.79
IGFBP5	1.96173E-73	2744.3
RPL9P17	1.64548E-10	2534.71
CADPS	8.40463E-08	2464.27
PRKG2	2.66448E-52	2406.66
LOC102725035	0.00179518	2315.28
SOX2	5.8545E-73	2298.42
NYNRIN	2.81821E-76	2295.13
LOC105379263	1.58835E-09	2258.33
MAP1LC3A	4.8342E-62	2055.8
EFS	4.21645E-78	2018.88
C9orf72	3.72144E-50	1934.96
SLC25A24P2	1.49744E-34	1897.25
MMP2	1.34523E-95	1874.42
NID1	3.3667E-103	1871.08
CYP1B1	2.05222E-86	1783.8
LOC105374103	2.24467E-11	1780.64
PTPN20	7.47638E-41	1649.22
PQLC2L	2.01363E-68	1527.47
KIT	1.0747E-85	1505.21

NFATC4	8.33019E-24	1497.6
THNSL2	2.1398E-112	1485.59
AEBP1	7.61137E-79	1446.32
LINC00906	1.9368E-111	1414.4
HIST1H3I	2.25968E-34	1393.49
PAK6	0.00120837	1254.94
CDH11	4.03828E-63	1227.81
TSPYL5	3.44031E-95	1223.21
MAL	6.50729E-81	1191.36
NPM2	9.32079E-71	1184.81
PCDH11X	3.4077E-83	1170.79
CCL18	1.32957E-68	1115.29
CPXM2	2.59825E-63	1081.93
PLXDC2	6.24004E-22	1071.62
PXDN	2.5552E-107	1036.54
CTAGE4	1.16203E-08	1022.91
LOC105370202	1.13966E-46	1020.79
SOX1	2.50359E-60	1019.78
LOC107984494	6.95499E-45	1009.92
OR2A20P	9.31835E-19	959.841
IFI44	3.22499E-69	937.1
LOC105373591	9.66923E-59	897.384
ZNF486	3.7739E-14	833.486
ZFP3	1.1826E-46	829.123
CD36	2.45564E-42	820.556
CCDC80	2.10327E-85	811.45
LOC107984523	5.35818E-87	788.845
PLEKHA6	7.19682E-88	783.903
RAMP1	1.69558E-74	778.853
MAGI2-AS3	5.80573E-59	778.648
ANXA1	1.04229E-53	755.356
AKR1C2	2.71185E-06	755.03
SERPINB1	1.15733E-25	734.033
LOC151174	2.40184E-59	720.766
LOC105376004	2.68197E-33	713.618
ITGA4	2.03082E-19	712.889
IRAK3	1.10154E-91	702.018
SLC16A2	5.88936E-80	693.419
CEMP1	6.5294E-07	656.751
ENPP2	9.86062E-82	655.978
TSPOAP1	2.55331E-57	644.567
C1QTNF5	2.43309E-70	634.774
PCDHGC3	4.80833E-58	628.464
PDK2	5.96949E-70	618.234
NBPF6	3.582E-62	615.304
ZSCAN12	3.62109E-63	606.394
MCHR1	7.76893E-47	584.558
ARHGEF5	3.07892E-16	569.273
LINGO2	1.17139E-47	553.746
SLC16A5	5.4939E-36	550.383
CDH2	3.51997E-77	549.835

Table shows bottom 100 downregulated genes comparing Melanocytes and SKMEL5 in 10%FBS

Gene Symbol	p-value	Fold Change
PHYHIPL	0.00268739	-118.376
IGFBP2	7.96149E-23	-119.331
KIF26A	5.20644E-26	-120.17
TEKT4	3.67286E-09	-123.212
ITK	1.24638E-17	-124.136
RPL21P40	5.34831E-08	-124.171
LOC100131536	0.00175397	-124.692
CHRD	8.56192E-25	-125.444
ZIC5	1.51986E-18	-126.27
LINC00052	8.15291E-28	-133.209
LOC101927098	8.95685E-20	-134.002
MIR9-3HG	1.97268E-20	-135.249
SYTL5	6.15617E-13	-136.611
ANKRD62	2.21824E-10	-136.833
RNVU1-8	1.70983E-11	-144.352
FNDC9	9.02677E-05	-145.27
LOC102724938	1.88398E-23	-149.542
UGT3A2	3.56216E-19	-150.646
SYT17	1.75014E-09	-154.37
FAM95A	2.59682E-09	-159.378
NKD2	0.000121105	-162.94
KIAA1462	1.9567E-23	-163.553
LOC107986170	9.28379E-20	-167.589
LOC105377685	1.95805E-17	-168.267
LOC105373057	8.46576E-13	-169.189
ASNSP1	3.74269E-34	-174.329
OR51B2	1.35274E-21	-175.861
C10orf126	4.41819E-10	-176.159
CLGN	1.60433E-17	-176.534
NECAP1P1	1.44917E-13	-178.51
METTL7B	1.48622E-18	-192.655
CCND2	0.000223077	-193.915
CFAP58-AS1	2.85444E-27	-202.859
C17orf77	4.65368E-10	-206.023
ATP6V0A4	1.46011E-18	-207.879
HMX1	1.02234E-18	-212.994
ADRB1	1.37173E-12	-214.408
MNX1-AS1	1.42313E-16	-221.522
KCNQ1	2.67877E-23	-231.639
NF1P6	1.23738E-17	-243.709
NEK2P2	1.3093E-14	-254.514
KIRREL2	2.81144E-12	-259.218
TM4SF19	2.11747E-14	-274.072
SV2B	5.38727E-19	-292.234
ADD2	3.45529E-31	-293.91
LOC101928797	2.00235E-17	-311.685
LOC100533706	3.93051E-25	-325.243
RIMS4	4.0239E-30	-328.501

VEGFC	3.98426E-12	-334.43
LOC107987036	7.65776E-08	-336.672
CKMT1B	1.95091E-34	-339.361
EPHB2	1.59605E-19	-344.144
XAGE1B	4.80324E-15	-347.655
LOC105375409	9.18314E-12	-352.467
LOC107987160	1.07656E-24	-352.705
SNORD116-18	0.00235945	-366.549
OR51B4	2.20124E-30	-383.684
OVAAL	2.14516E-28	-388.92
ZIC3	3.03907E-16	-392.531
LOC400655	7.62958E-16	-397.181
ADAMTSL2	5.06787E-20	-400.801
TESC	5.73433E-12	-410.383
ALDH2	1.75683E-28	-458.217
IGFBPL1	8.5648E-27	-459.489
LOC107985517	2.05136E-25	-467.429
LINC00479	1.3306E-21	-467.883
FAM167A	5.72388E-25	-468.267
IGFN1	1.55571E-21	-470.196
PRLR	6.00518E-16	-475.958
LVCAT1	2.2782E-17	-484.537
DLX6-AS1	2.11882E-15	-513.117
IL13RA2	1.27006E-20	-532.296
ATP1A3	3.66667E-23	-535.088
GCSAML	1.38572E-20	-555.817
NOTUM	7.16214E-26	-572.533
CHGB	1.33283E-06	-574.767
ASB4	1.75229E-17	-801.608
SYT13	4.50485E-22	-906.806
LOC100506688	6.59475E-28	-1002.2
MAT1A	2.29369E-30	-1044.27
ELMOD1	1.20378E-24	-1048.79
MECOM	6.85187E-25	-1157.78
SEL1L3	1.34601E-24	-1258.56
VCX2	2.82629E-05	-1330.58
DNER	3.95308E-27	-1546.04
FERMT1	7.06929E-27	-2306.75
MATK	3.60241E-18	-2959.68
SNORA7A	2.45298E-11	-3088.79
HBE1	8.19093E-27	-3728.45
GAGE1	3.81735E-20	-3871.4
PASD1	5.75755E-26	-4473.11
MAGEC2	1.41643E-22	-6009.1
GAGE2A	2.25963E-19	-7037.93
LOC105371108	1.27436E-05	-19848.7
LOC100419003	8.76272E-08	-746523
LOC101928058	0.000415822	-2815810
LOC107987324	9.75752E-10	-281716000
POTEG	4.4282E-10	-11450800000
LINC00668	1.01987E-13	-15330900000
LOC390956	0.00202244	-57088800000

Table shows top 100 upregulated genes comparing Melanocytes and SKMEL5 in 1%FBS

Gene Symbol	p-value	Fold Change
IGKV1OR9-2	4.64608E-07	5878790000
DNM1P40	0.00095982	1173750000
LOC107986995	0.000280557	500366000
LIMS3-LOC440895	0.00110661	6048280
SERHL	0.00111652	3118770
NBPF4	3.70452E-27	981307
AQP7P5	1.25989E-33	842619
CTAGE8	6.14815E-06	789450
LOC107987069	0.00296203	382671
GSTT2	5.31061E-47	297136
AQP7P4	4.55868E-08	44804.5
MTAP	4.55319E-41	24971.1
PCDH11Y	4.10799E-07	22856.3
IFI44L	1.18292E-69	13369.9
THNSL2	2.5033E-111	9761.72
LOC101928605	3.2798E-11	7441.92
EVC2	4.56324E-73	7290.48
MTHFD1P1	8.12053E-05	5981.76
ARHGEF35	1.49719E-05	5773.92
ZSCAN12	2.68069E-62	5544.91
TGFB1I1	1.13053E-63	4436.56
MAP1LC3A	4.14776E-61	4129.97
SECTM1	2.41988E-73	4068.3
MMP2	1.54304E-94	3523.87
ZFP3	7.28563E-46	2923.43
KIT	1.15217E-84	2795.06
IRAK3	1.10247E-90	2492.79
FXYD3	1.14965E-54	2483.52
FZD10	7.08392E-69	2399.37
CAAP1	6.0167E-33	2290.42
CEMP1	9.38743E-07	2255.94
IFI27	4.04218E-61	2247.57
PLXDC2	1.94024E-21	2037.66
CPQ	2.26459E-53	1972.61
IGFBP5	2.03399E-72	1861.65
LOC105369482	3.7353E-38	1850.35
CMBL	9.4105E-18	1698.45
PQLC2L	1.91513E-67	1696.73
FZD10-AS1	1.12864E-91	1642.19
CDKN2A	0.000137857	1605.3
COL11A1	3.14082E-85	1604.25
IRF6	1.09801E-32	1556.08
LINC00906	2.4425E-110	1551.71
XIST	5.264E-100	1488.13
CYP1B1	2.37781E-85	1331.21
SLC9A7P1	2.46677E-34	1240.98
CDH3	3.56399E-91	1226.69
LOC400043	5.12347E-30	1190.12
MOB3B	8.2352E-99	1187.21

LOC107984523	5.71302E-86	1157.1
TMEM132D	5.8662E-104	1150.84
CES4A	3.20852E-36	1130.19
CLEC18B	0.00120106	1116.88
RPS3AP35	4.53856E-65	1096.76
PCDHGC3	3.68779E-57	1090.92
LOC401296	8.56648E-40	1074.73
GALNT5	2.11434E-38	1066.44
ANXA1	7.74581E-53	1062.6
SLC25A24P1	7.53656E-71	1033.7
LOC107985742	8.99685E-73	1014.59
PPP1R3C	1.25745E-79	1006.57
PRKG2	2.18263E-51	973.528
PRDX2	1.39018E-80	964.03
COL1A2	9.4819E-107	947.428
SLFN11	7.84582E-84	928.997
ENPP2	1.01128E-80	901.078
ARHGEF5	7.38497E-16	847.762
TRIM61	1.71135E-44	835.272
GPR173	4.48167E-86	827.449
TMEM154	2.55326E-33	826.309
CHST3	1.0211E-103	817.782
RPP25	2.67188E-29	781.046
LOC107986304	5.79191E-26	754.764
EPHX2	7.67752E-84	750.534
SERPING1	1.01863E-24	739.674
LOC107986990	2.14437E-58	729.303
ADAMTSL1	3.57178E-84	724.388
NKAPL	4.18073E-34	717.473
INSC	4.93617E-36	704.385
ZNF662	2.11748E-57	701.028
PDK2	5.68854E-69	692.957
LOC107984494	4.79786E-44	692.059
PLAA	2.26068E-59	691.87
SERPINH1	4.52441E-81	682.771
KIAA1755	9.53995E-61	682.122
MAL	7.73262E-80	670.476
LINC00639	1.49511E-25	656.625
CCL2	6.04437E-54	652.088
PCDHGC4	4.21722E-26	651.937
TDRD9	1.72874E-11	651.475
ZNF334	2.13546E-38	650.625
TP73-AS1	2.29353E-17	646.515
OR2A20P	2.68775E-18	630.37
MAGI2-AS3	5.12497E-58	626.543
SGIP1	2.24158E-75	625.275
LOC105373592	5.18531E-79	623.476
SLC16A2	6.48283E-79	618.401
CDH11	4.03958E-62	617.341
LOC105372698	3.26078E-75	588.992
CADM3	3.48207E-87	588.415

Table shows bottom 100 downregulated genes comparing Melanocytes and SKMEL5 in 1%FBS

Gene Symbol	p-value	Fold Change
ILDR2	6.99115E-52	-195.169
INHBA	1.0969E-37	-198.687
LOC101929745	0.0033878	-199.022
CCND2	0.00017273	-200.711
VCX	3.43619E-41	-203.105
LOC105378456	5.00598E-12	-205.671
HOXD1	5.94636E-56	-206.177
IGFBP2	7.12275E-45	-206.945
LOC105370850	2.07565E-43	-207.373
IGFN1	3.28853E-06	-208.738
DLK1	3.32585E-23	-211.392
KIAA1462	2.42025E-32	-211.619
NKD2	4.75437E-07	-220.358
NF1P6	1.20384E-15	-228.565
NEK2P2	3.50009E-12	-229.527
ZIC5	3.5496E-40	-229.767
KCNQ1	4.24926E-23	-234.032
LOC102724938	1.98845E-42	-242.177
PHF21B	4.82058E-05	-249.51
MNX1-AS1	1.55606E-19	-251.908
RPL21P40	1.20478E-22	-252.657
LOC105373553	0.00302393	-260.803
ANKRD62	1.30302E-26	-264.023
ADD2	7.54359E-27	-266.63
CALB2	7.65732E-19	-267.65
ITK	8.13173E-47	-267.76
LOC107986170	1.5837E-36	-269.702
ZIC3	3.42654E-09	-274.231
LOC100533706	1.54864E-19	-276.529
CRLF2	4.40653E-19	-282.371
SCN3B	1.43111E-43	-290.101
SYT17	1.67837E-24	-297.524
OR51B4	8.13006E-23	-315.474
OGDHL	6.94944E-55	-317.503
TERT	0.00103211	-329.794
HMX1	2.95947E-33	-329.94
RIMS4	3.76E-32	-351.344
LOC102723733	1.51926E-27	-357.195
OVAAL	7.65717E-26	-368.553
PADI2	0.0023909	-370.903
SV2B	5.80751E-27	-381.487
ADRB1	1.84846E-29	-394.101
KIRREL2	1.5336E-22	-397.81
SYTL5	9.22516E-54	-413.636
C10orf126	1.35595E-36	-444.651
IL13RA2	5.20704E-16	-455.022
CKMT1B	1.74218E-47	-455.187
LOC107987160	7.36054E-35	-466.888

ALDH2	2.58061E-29	-476.166
DLX6-AS1	1.65359E-13	-478.103
NOTUM	6.63598E-20	-482.623
ADAMTSL2	1.28585E-26	-499.784
LOC107985517	1.61165E-27	-504.083
ATP6V0A4	5.10221E-55	-506.409
FAM167A	2.38149E-28	-523.15
IGFBPL1	1.6266E-33	-555.018
LOC400655	2.40387E-25	-560.436
C17orf77	8.45437E-41	-571.205
ATP1A3	2.0732E-26	-598.584
LMX1B	3.77168E-05	-601.334
EPHB2	7.15768E-41	-614.801
XAGE1B	8.40508E-34	-632.93
PHYHIPL	3.62971E-38	-672.306
TM4SF19	6.13658E-49	-707.613
LOC107987036	3.03593E-24	-730.807
LVCAT1	1.66707E-30	-738.822
TESC	1.89587E-30	-800.747
GCSAML	4.24128E-33	-803.819
LOC100506688	3.68178E-22	-860.881
VEGFC	1.5744E-44	-887.137
LOC100131536	1.29229E-54	-941.767
LINC00479	1.5309E-57	-1075.07
LOC105375409	1.54892E-51	-1093.34
MAT1A	2.64189E-32	-1114.26
ELMOD1	3.27853E-32	-1305.71
MECOM	2.79042E-32	-1433.53
EPHA7	4.22874E-14	-1459.98
CXCR4	0.000349466	-1519.99
SYT13	1.68468E-46	-1668.94
LOC107985902	3.60727E-05	-1775.92
RPL7P50	0.000475394	-1792.69
CHGB	4.93315E-35	-1856.06
ASB4	9.55139E-54	-1991.7
PRLR	1.11954E-81	-2019.45
SEL1L3	5.43057E-46	-2121.55
DNER	7.2892E-48	-2515.54
FERMT1	4.18713E-31	-2621.47
HBE1	2.33401E-22	-3320.78
VCX2	1.08762E-26	-4056.45
MATK	9.68272E-39	-5312.8
PASD1	2.24808E-38	-6183.06
GAGE1	6.10292E-59	-9589.21
MAGEC2	1.27438E-48	-11349.5
LOC105371108	0.00374277	-13181.1
GAGE2A	2.81797E-59	-18020.9
LOC101928058	0.000430993	-2854740
LOC107987324	0.000250783	-165096000
POTEG	1.7017E-15	-15465600000
LINC00668	6.44161E-23	-22155400000
LOC390956	6.5694E-07	-95796100000

Table shows top 100 upregulated genes comparing SKMEL5 in 10% FBS (AD) and SKMEL5 in 1%FBS (SUS)

Gene Symbol	p-value	Fold Change
SNORA17A	2.7095E-08	2.97165E+12
SNORA7A	5.08115E-10	126685
SNORA14B	8.61067E-09	157.592
CD200R1	0.00129384	76.4729
KRT81	0.000665177	56.2075
LOC102723994	1.25949E-06	55.4918
DEFB122	0.000103352	46.2067
LOC650866	3.34521E-05	42.7203
RNA5SP71	0.00107955	38.7394
RNA5SP338	0.000442413	37.0023
MAPK6PS2	8.85235E-07	34.8134
RN7SL555P	0.000101892	32.428
SNORA50A	1.19787E-10	31.6699
CES1P2	0.00026572	29.9016
SNORA17B	1.23048E-17	27.6418
VN1R81P	0.00135387	25.2872
RNU5A-1	1.53049E-12	24.9127
LOC317727	0.00144495	24.2338
RNVU1-1	9.79181E-05	22.8012
RNY5	4.56333E-10	21.9286
RNA5SP215	4.66521E-08	20.3167
LOC105373485	0.000453013	20.2405
RN7SL674P	8.4097E-10	20.0567
SNORD99	0.00187499	18.2261
LOC107987148	1.77281E-05	17.0026
TRNM	5.17006E-11	15.8843
ATP6V1G1P7	1.85095E-06	15.6338
RNA5SP374	1.07854E-06	15.0114
RNU5E-6P	1.68839E-10	14.8124
SCARNA1	0.000028676	14.7475
TRNS2	2.66052E-09	14.571
SNORD100	0.000242803	13.1673
RNU12	5.71792E-12	12.9174
GABRG1	1.77377E-08	12.8844
RN7SKP220	0.000991327	12.6182
RNU4ATAC	8.3274E-12	12.4788
SNAR-A1	2.28705E-05	12.0752
SNAR-A2	2.28705E-05	12.0752
LOC100506885	2.92763E-09	11.9484
LOC107987279	0.00113216	11.9358
SNORA72	4.22545E-09	11.7934
RNU2-2P	3.97356E-11	11.783
SNORA44	3.29314E-10	11.7769
SNORA65	3.406E-08	11.5245
SNORA3A	5.45634E-11	11.1581
SNORD67	4.67918E-06	10.8115
SNORA26	7.91627E-06	10.678
RNU5B-1	9.64405E-13	10.659

SNORA38B	2.36009E-10	10.5679
SC4MOP	0.000178178	10.5553
SNORA38	3.39796E-13	10.4372
SNORA75	0.000207083	10.3891
MTRNR2L1	0.000151923	10.2672
SGCG	1.91214E-07	10.161
RNU4-1	4.45278E-15	10.1343
MTRNR2L12	5.50155E-05	10.0592
SNORA46	9.36309E-05	9.60471
TRNV	8.58647E-11	9.53455
RN7SL3	1.46419E-05	9.53184
SFR1P1	0.00182594	9.41791
RN7SL546P	1.70231E-05	9.39921
RNU5E-1	5.75087E-08	9.39405
KRT86	5.09254E-08	9.22872
LOC105374879	0.000248594	9.1428
SNORA9	1.75527E-11	9.14051
LOC105379634	0.000837911	9.12302
RNU6ATAC	4.98772E-16	8.93863
LOC105378951	0.000884446	8.90589
RNU4-2	1.51825E-13	8.90085
LOC105372826	0.000798547	8.87786
SNORA66	2.94827E-08	8.7176
SNORA50C	5.94601E-13	8.67295
TRNH	0.000342853	8.55314
SNORA74B	2.207E-09	8.52399
IMPDH1P3	0.000210668	8.26768
LOC101927785	1.33687E-06	8.0836
RN7SL128P	4.77043E-14	8.06908
LOC390314	0.000190358	8.06017
SNORA13	0.000169453	8.00787
SNORA61	1.10232E-14	7.90061
LOC105376934	1.99605E-05	7.8763
RPPH1	4.06103E-16	7.79263
RNU5D-1	0.000202259	7.77612
MTRNR2L2	0.00073657	7.70575
SNORA2A	1.39487E-09	7.69167
SNORA57	3.22545E-12	7.63851
SNORA37	0.000021581	7.6128
SNORA23	5.07219E-11	7.60409
PAPPA-AS1	0.000118019	7.59859
TRNY	1.26582E-13	7.56982
LOC100507537	0.000578417	7.55205
SCARNA10	1.26523E-15	7.45859
SNORA1	0.000754278	7.44988
SNORA25	0.000019362	7.38044
SNORA62	2.52197E-11	7.35492
SNORA80B	1.5324E-15	7.26126
LRRC3B	1.83488E-11	7.22631
ASTN2-AS1	8.00838E-05	7.21246
FIBIN	6.78994E-05	7.02612
LOC100419003	1.71718E-05	7.0142

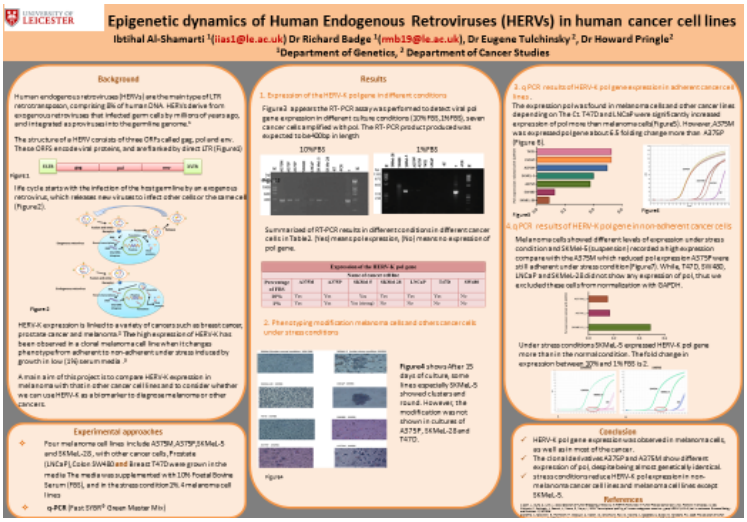
Table shows bottom 100 downregulated genes comparing SKMEL5 in 10% FBS (AD) and SKMEL5 in 1%FBS (SUS)

Gene Symbol	p-value	Fold Change
FFAR1	6.7396E-06	-13.9989
CACNA2D3	0.000313574	-14.0518
DAP3P1	0.000409618	-14.102
SNORD136	0.0014527	-14.1033
CFB	6.50799E-08	-14.156
LOC105378709	0.00139027	-14.1971
LOC100419668	7.33696E-05	-14.242
HSPE1P26	4.62398E-09	-14.3417
LOC102724591	2.15881E-05	-14.4588
LINC01617	2.50595E-12	-14.6775
CSMD3	1.11309E-10	-14.6874
CNGA1	3.14407E-11	-14.8162
LOC107986517	3.94535E-05	-14.9987
LOC107984365	2.71881E-06	-15.0598
ATP1A4	0.000500922	-15.3279
CFAP47	5.08912E-10	-15.3331
LOC105376713	0.000580044	-15.8198
LOC101928526	0.000255955	-15.8259
RPLP0P5	0.000600802	-16.0262
LACAT1	0.00034129	-16.1141
RPL31P58	0.000061596	-16.1409
PPP1R27	0.000226533	-16.3185
LOC105375422	0.000235206	-16.4513
NLRP4	3.4518E-20	-16.7034
LINC01606	2.12076E-12	-17.0768
LOC284241	0.00173027	-17.4788
LOC439933	0.000378493	-17.4867
TRIM80P	0.00161842	-17.4965
SLC25A41	1.1755E-09	-17.5321
LOC105371893	0.00106743	-17.8346
LOC105377762	1.0816E-08	-18.1471
LOC101927609	0.00100558	-18.5081
EHF	1.22094E-09	-18.6312
LOC105375871	0.000072611	-18.7778
LOC107985528	0.000247705	-18.8817
RN7SL213P	0.00082757	-19.0585
CXCR4	0.00140829	-19.3053
LOC105378145	0.000067866	-19.3657
LOC102723339	4.35443E-06	-19.3846
IL10	0.000495767	-19.7659
OVOL3	0.00126566	-19.988
LOC646241	4.9586E-21	-20.2905
C5orf47	1.18352E-15	-20.4012
LOC442293	2.82582E-05	-20.4801
LOC101929099	8.34669E-08	-21.5352
RPS12P31	0.000412319	-21.5474
PRDX3P2	0.000545723	-21.5743
RPL7P44	9.86108E-08	-22.7073

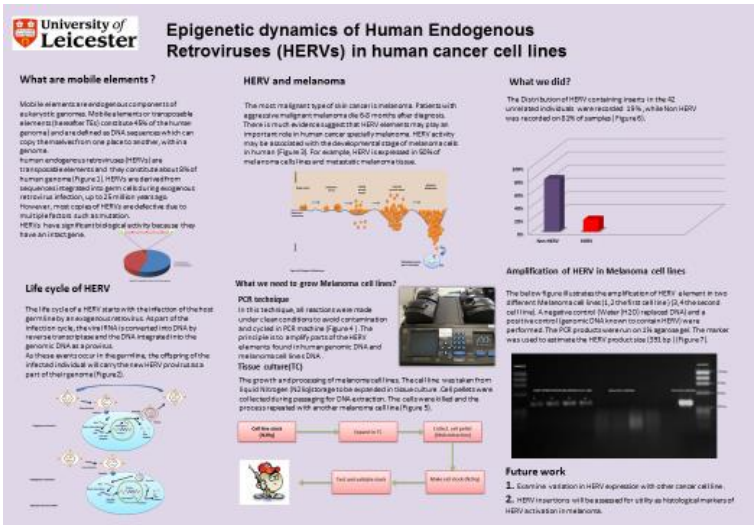
LOC105371636	3.9711E-07	-23.4651
HMGB3P6	0.000855279	-23.6491
CGA	1.34161E-08	-24.1408
LOC105370429	0.000192486	-24.2667
LOC101928423	8.97686E-08	-24.3188
LOC101926994	0.000919437	-24.5723
BTC	1.98849E-09	-25.0355
SLC23A1	6.4781E-07	-25.5204
PENK	2.27113E-07	-26.3705
LUCAT1	0.00145436	-26.6765
PMM2P2	0.00183975	-26.7331
SPDYE7P	0.000644177	-27.3549
DIRC3-AS1	4.21259E-05	-27.8721
DUX4L19	0.000106788	-28.3146
LOC105373689	7.54964E-06	-28.8302
IMMP1LP3	0.00099829	-29.1929
LOC401286	0.000234958	-29.2671
RPL39P14	8.89386E-05	-29.2923
PTCHD1-AS	1.29975E-09	-30.4369
PCDHA1	3.28411E-07	-31.1496
EVPL	1.43142E-20	-31.3096
SRIP3	0.0011404	-31.4132
IL7R	0.000477304	-31.9731
LOC107986851	7.30631E-05	-33.7208
LOC101928032	1.68506E-07	-34.891
DNM1P38	0.000922145	-35.6972
SOSTDC1	0.00191158	-35.935
LINC00304	0.000529441	-36.4615
LOC101929440	0.000177265	-36.746
RN7SL678P	6.26814E-05	-37.14
LOC105377068	0.000167306	-39.2401
MISP	2.54242E-10	-40.1917
INHBA	5.40932E-34	-40.5541
PSG4	0.000159559	-41.8329
OR7E47P	1.00921E-05	-41.8372
LOC105375539	3.90077E-09	-43.5809
LOC105369304	2.71103E-06	-47.0727
LOC105369594	0.000215441	-47.5228
LOC105370780	4.70814E-05	-47.7751
CEACAMP10	1.31001E-07	-49.3842
LOC107986421	0.00122372	-54.7462
PSG9	3.15409E-08	-55.597
TBX4	6.69905E-09	-56.1233
ATP5L2	0.000159626	-58.0974
RPS3AP2	1.83015E-06	-65.5006
RPL39P40	2.40016E-07	-65.7169
LOC105374738	5.61052E-09	-66.7985
LOC100507634	1.85197E-05	-75.249
FGF9	6.28017E-14	-80.8744
LOC105369379	5.11725E-15	-139.797
GP1BB	0.000555712	-144.087
RPL7P50	0.00102591	-2905.61

Participating conferences

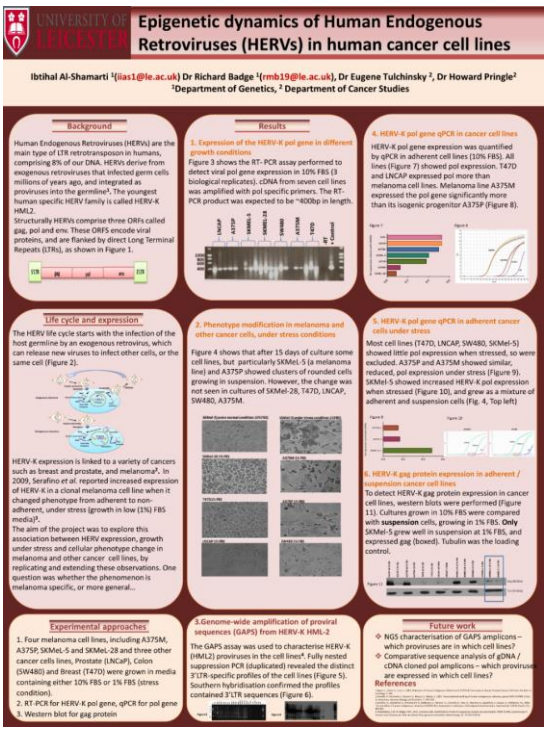
PGR seminar day /University of Leicester 2015



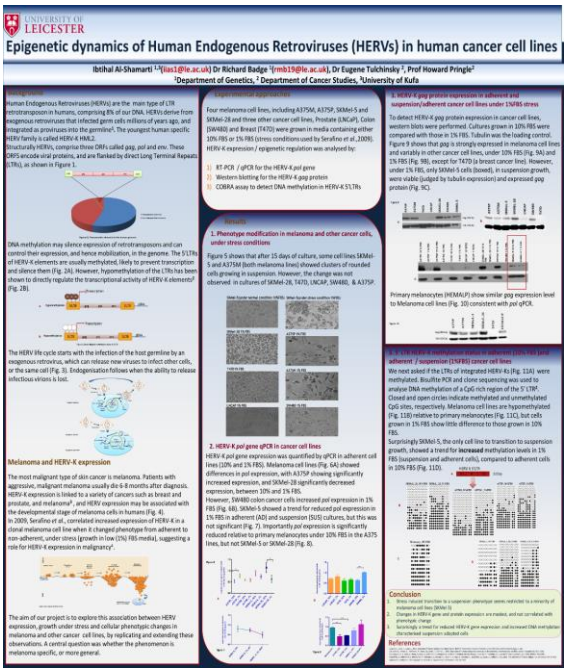
Cancer studies development day and Festival postgraduate research /University of Leicester 2015



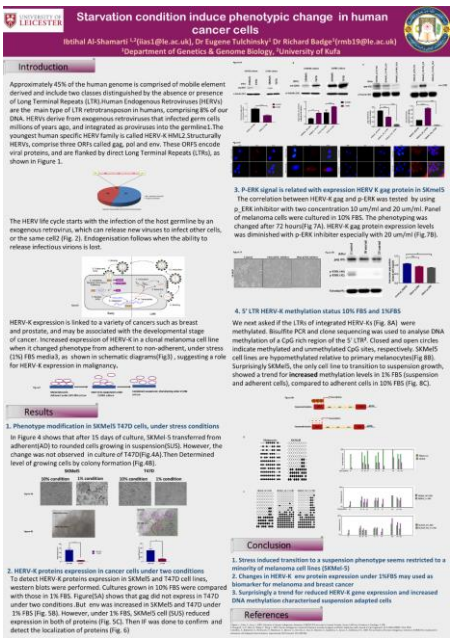
The third international conference on Transposable Elements conference (ICTE)/ France 2016



Transposable Elements Conference /USA 2016



BACR Development of Cancer Medicines _ Royal Society of Medicines London 2017



Starvation condition induce phenotypic change in human cancer cells

Ibtihal Al-Shamarti [1,2](ias1@le.ac.uk), Dr Eugene Tulchinsky[1] Dr Richard Badge[1](rmb19@le.ac.uk)

[1]Department of Genetics & Genome Biology, [2]University of Kufa

Introduction

Approximately 45% of the human genome is comprised of mobile element derived and include two classes distinguished by the absence or presence of Long Terminal Repeats (LTR).Human Endogenous Retroviruses (HERVs) are the main type of LTR retrotransposon in humans, comprising 8% of our DNA. HERVs derive from exogenous retroviruses that infected germ cells millions of years ago, and integrated as proviruses into the germline1.The youngest human specific HERV family is called HERV-K HML2.Structurally HERVs, comprise three ORFs called gag, pol and env. These ORFS encode viral proteins, and are flanked by direct Long Terminal Repeats (LTRs), as shown in Figure 1.

The HERV life cycle starts with the infection of the host germline by an exogenous retrovirus, which can release new viruses to infect other cells, or the same cell2 (Fig. 2). Endogenisation follows when the ability to release infectious virions is lost.

HERV-K expression is linked to a variety of cancers such as breast and prostate, and may be associated with the developmental stage of cancer. Increased expression of HERV-K in a clonal melanoma cell line when it changed phenotype from adherent to non-adherent, under stress (1%) FBS media3, as shown in schematic diagrams(Fig3) , suggesting a role for HERV-K expression in malignancy.

Results

1. Phenotype modification in SKMel5 T47D cells, under stress conditions

In Figure 4 shows that after 15 days of culture, SKMel-5 transferred from adherent(AD) to rounded cells growing in suspension(SUS). However, the change was not observed in culture of T47D(Fig.4A).Then Determined level of growing cells by colony formation (Fig.4B).

2. HERV-K proteins expression in cancer cells under two conditions

To detect HERV-K proteins expression in SKMel5 and T47D cell lines, western blots were performed. Cultures grown in 10% FBS were compared with those in 1% FBS. Figure(5A) shows that gag did not express in T47D under two conditions .But env was increased in SKMel5 and T47D under 1% FBS (Fig. 5B). However, under 1% FBS, SKMel5 cell (SUS) reduced expression in both of proteins (Fig. 5C). Then IF was done to confirm and detect the localization of proteins (Fig. 6)

3. P-ERK signal is related with expression HERV K gag protein in SKmelS

The correlation between HERV-K gag and p-ERK was tested by using p_ERK inhibitor with two concentration 10 um/ml and 20 um/ml. Panel of melanoma cells were cultured in 10% FBS. The phenotyping was changed after 72 hours(Fig 7A). HERV-K gag protein expression levels was diminished with p-ERK inhibiter especially with 20 um/ml (Fig.7B).

4. 5' LTR HERV-K methylation status 10% FBS and 1%FBS

We next asked if the LTRs of integrated HERV-Ks (Fig. 8A) were methylated. Bisulfite PCR and clone sequencing was used to analyse DNA methylation of a CpG rich region of the 5' LTR[4]. Closed and open circles indicate methylated and unmethylated CpG sites, respectively. SKMel5 cell lines are hypomethylated relative to primary melanocytes(Fig 8B). Surprisingly SKMel5, the only cell line to transition to suspension growth, showed a trend for **increased** methylation levels in 1% FBS (suspension and adherent cells), compared to adherent cells in 10% FBS (Fig. 8C).

Conclusion

1. Stress induced transition to a suspension phenotype seems restricted to a minority of melanoma cell lines (SKMel-5)
2. Changes in HERV-K env protein expression under 1%FBS may used as biomarker for melanoma and breast cancer
3. Surprisingly a trend for reduced HERV-K gene expression and increased DNA methylation characterised suspension adapted cells

References

Seed and Soil: In Vivo Models of Metastasis certificate _Germany, Berlin 2017

Certificate of Attendance

This is to certify that the below mentioned person attended the 1st joint EACR-MRS Conference on Seed and Soil: In Vivo Models of Metastasis, held in Berlin, Germany, 27 - 29 November 2017.

IBTIHAL AL-SHAMARTI

United kingdom

On behalf of the Scientific Programme Committee:

Janine Terra Erler (Denmark)
Jonathan Sleeman (Germany)
Ulrike Stein (Germany)
Christof von Kalle (Germany)

EACR
European Association for Cancer Research

9 References

Agoni, L., Guha, C., Lenz, J., 2013a. Detection of Human Endogenous Retrovirus K (HERV-K) Transcripts in Human Prostate Cancer Cell Lines. *Frontiers in Oncology.* **3,** 180.

Alaghband, Y., Bredy, T.W., Wood, M.A., 2016. The role of active DNA demethylation and Tet enzyme function in memory formation and cocaine action. *Neuroscience Letters.* **625,** 40-46.

Antequera, F. & Bird, A., 1993. Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America.* **90,** 11995-11999.

Argaw-Denboba, A., Balestrieri, E., Serafino, A., Cipriani, C., Bucci, I., Sorrentino, R., Sciamanna, I., Gambacurta, A., Sinibaldi-Vallebona, P., Matteucci, C., 2017. HERV-K activation is strictly required to sustain CD133+ melanoma cells with stemness features. *Journal of Experimental & Clinical Cancer Research: CR.* **36,** 20-016-0485-x.

Ballestar, E., Paz, M.F., Valle, L., Wei, S., Fraga, M.F., Espada, J., Cigudosa, J.C., Huang, T.H., Esteller, M., 2003. Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. *The EMBO Journal.* **22,** 6335-6345.

Bannert, N. & Kurth, R., 2006. The evolutionary dynamics of human endogenous retroviral families. *Annual Review of Genomics and Human Genetics.* **7,** 149-173.

Baylin, S.B., Esteller, M., Rountree, M.R., Bachman, K.E., Schuebel, K., Herman, J.G., 2001. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Human Molecular Genetics.* **10,** 687-692.

Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., Moran, J.V., 2010. LINE-1 retrotransposition activity in human genomes. *Cell.* **141,** 1159-1170.

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. *Proceedings of the National Academy of Sciences of the United States of America.* **101,** 4894-4899.

Benesova, M., Trejbalova, K., Kovarova, D., Vernerova, Z., Hron, T., Kucerova, D., Hejnar, J., 2017. DNA hypomethylation and aberrant expression of the human endogenous retrovirus ERVWE1/syncytin-1 in seminomas. *Retrovirology.* **14,** 20-017-0342-9.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., Schreiber, S.L., Lander, E.S., 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell.* **120,** 169-181.

Best, S., Le Tissier, P.R., Stoye, J.P., 1997. Endogenous retroviruses and the evolution of resistance to retroviral infection. *Trends in Microbiology.* **5,** 313-318.

Bestor, T.H., 2000. The DNA methyltransferases of mammals. *Human Molecular Genetics.* **9,** 2395-2402.

Bhardwaj, N., Montesion, M., Roy, F., Coffin, J.M., 2015. Differential expression of HERV-K (HML-2) proviruses in cells and virions of the teratocarcinoma cell line Tera-1. *Viruses.* **7,** 939-968.

Bieda, K., Hoffmann, A., Boller, K., 2001. Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines. *The Journal of General Virology.* **82,** 591-596.

Bollag, G., Hirth, P., Tsai, J., Zhang, J., Ibrahim, P.N., Cho, H., Spevak, W., Zhang, C., Zhang, Y., Habets, G., Burton, E.A., Wong, B., Tsang, G., West, B.L., Powell, B., Shellooe, R., Marimuthu, A., Nguyen, H., Zhang, K.Y., Artis, D.R., Schlessinger, J., Su, F., Higgins, B., Iyer, R., D'Andrea, K., Koehler, A., Stumm, M., Lin, P.S., Lee, R.J., Grippo, J., Puzanov, I., Kim, K.B., Ribas, A., McArthur, G.A., Sosman, J.A., Chapman, P.B., Flaherty, K.T., Xu, X., Nathanson, K.L., Nolop, K., 2010. Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature.* **467,** 596-599.

Bonnelykke-Behrndtz, M.L., Steiniche, T., Norgaard, P., Danielsen, A.V., Damsgaard, T.E., Christensen, I.J., Bastholt, L., Moller, H.J., Schmidt, H., 2017. Loss of E-cadherin as Part of a Migratory Phenotype in Melanoma Is Associated With Ulceration. *The American Journal of Dermatopathology.*

Box, N.F., Duffy, D.L., Chen, W., Stark, M., Martin, N.G., Sturm, R.A., Hayward, N.K., 2001. MC1R genotype modifies risk of melanoma in families segregating CDKN2A mutations. *American Journal of Human Genetics.* **69,** 765-773.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., Kazazian, H.H., Jr, 2003. Hot L1s account for the bulk of retrotransposition in the human population. *Proceedings of the National Academy of Sciences of the United States of America.* **100,** 5280-5285.

Bu, Y. & Diehl, J.A., 2016. PERK Integrates Oncogenic Signaling and Cell Survival During Cancer Development. *Journal of Cellular Physiology.* **231,** 2088-2096.

Buscher, K., Hahn, S., Hofmann, M., Trefzer, U., Ozel, M., Sterry, W., Lower, J., Lower, R., Kurth, R., Denner, J., 2006. Expression of the human endogenous retrovirus-K transmembrane envelope, Rec and Np9 proteins in melanomas and melanoma cell lines. *Melanoma Research.* **16,** 223-234.

Buscher, K., Trefzer, U., Hofmann, M., Sterry, W., Kurth, R., Denner, J., 2005. Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. *Cancer Research.* **65,** 4172-4180.

Calin, G.A. & Croce, C.M., 2006. MicroRNA signatures in human cancers. *Nature Reviews.Cancer.* **6,** 857-866.

Calvani, M., Pelon, F., Comito, G., Taddei, M.L., Moretti, S., Innocenti, S., Nassini, R., Gerlini, G., Borgognoni, L., Bambi, F., Giannoni, E., Filippi, L., Chiarugi, P., 2015. Norepinephrine promotes tumor microenvironment reactivity through beta3-adrenoreceptors during melanoma progression. *Oncotarget.* **6,** 4615-4632.

Cavener, D.R., Gupta, S., McGrath, B.C., 2010. PERK in beta cell biology and insulin biogenesis. *Trends in Endocrinology and Metabolism: TEM.* **21,** 714-721.

Cegolon, L., Salata, C., Weiderpass, E., Vineis, P., Palu, G., Mastrangelo, G., 2013. Human endogenous retroviruses and cancer prevention: evidence and prospects. *BMC Cancer.* **13,** 4-2407-13-4.

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 beta-catenin, ERK, Akt and Notch1 and promoting the growth of human leukemia stem/progenitor cells. *Leukemia.* **27,** 1469-1478.

Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N.S., Cope, L.M., Snyder, A., Makarov, V., Budhu, S., Slamon, D.J., Wolchok, J.D., Pardoll, D.M., Beckmann, M.W., Zahnow, C.A., Merghoub, T., Chan, T.A., Baylin, S.B., Strick, R., 2015. Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell.* **162,** 974-986.

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. *Journal of Virology.* **89,** 7187-7201.

de Parseval, N. & Heidmann, T., 2005. Human endogenous retroviruses: from infectious elements to human genes. *Cytogenetic and Genome Research.* **110,** 318-332.

Deininger, P.L., Moran, J.V., Batzer, M.A., Kazazian, H.H.,Jr, 2003. Mobile elements and mammalian genome evolution. *Current Opinion in Genetics & Development.* **13,** 651-658.

Denne, M., Sauter, M., Armbruester, V., Licht, J.D., Roemer, K., Mueller-Lantzsch, N., 2007. Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. *Journal of Virology.* **81,** 5607-5616.

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 Research.* **16,** 1548-1556.

Dunn, C.A., Medstrand, P., Mager, D.L., 2003. An endogenous retroviral long terminal repeat is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon. *Proceedings of the National Academy of Sciences of the United States of America.* **100,** 12841-12846.

Ehrlich, M., 2002. DNA hypomethylation, cancer, the immunodeficiency, centromeric region instability, facial anomalies syndrome and chromosomal rearrangements. *The Journal of Nutrition.* **132,** 2424S-2429S.

Esteller, M., Cordon-Cardo, C., Corn, P.G., Meltzer, S.J., Pohar, K.S., Watkins, D.N., Capella, G., Peinado, M.A., Matias-Guiu, X., Prat, J., Baylin, S.B., Herman, J.G., 2001. p14ARF silencing by promoter hypermethylation mediates abnormal intracellular localization of MDM2. *Cancer Research.* **61,** 2816-2821.

Feinberg, A.P. & Vogelstein, B., 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature.* **301,** 89-92.

Florl, A.R., Lower, R., Schmitz-Drager, B.J., Schulz, W.A., 1999. DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas. *British Journal of Cancer.* **80,** 1312-1321.

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. *Journal of Virology.* **85,** 3436-3448.

Gallo, R.C., Sarin, P.S., Gelmann, E.P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V.S., Mann, D., Sidhu, G.D., Stahl, R.E., Zolla-Pazner, S., Leibowitch, J., Popovic, M., 1983. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science (New York, N.Y.).* **220,** 865-867.

Gao, X., Havecker, E.R., Baranov, P.V., Atkins, J.F., Voytas, D.F., 2003. Translational recoding signals between gag and pol in diverse LTR retrotransposons. *RNA (New York, N.Y.).* **9,** 1422-1430.

Gifford, R. & Tristem, M., 2003. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes.* **26,** 291-315.

Goering, W., Schmitt, K., Dostert, M., Schaal, H., Deenen, R., Mayer, J., Schulz, W.A., 2015. Human endogenous retrovirus HERV-K(HML-2) activity in prostate cancer is dominated by a few loci. *The Prostate.* **75,** 1958-1971.

Gray-Schopfer, V., Wellbrock, C., Marais, R., 2007. Melanoma biology and new targeted therapy. *Nature.* **445,** 851-857.

Groudine, M., Eisenman, R., Weintraub, H., 1981. Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. *Nature.* **292,** 311-317.

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., Wysocka, J., 2015. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature.* **522,** 221-225.

Hancks, D.C. & Kazazian, H.H.,Jr, 2012. Active human retrotransposons: variation and disease. *Current Opinion in Genetics & Development.* **22,** 191-203.

Hanke, K., Chudak, C., Kurth, R., Bannert, N., 2013. The Rec protein of HERV-K(HML-2) upregulates androgen receptor activity by binding to the human small glutamine-rich tetratricopeptide repeat protein (hSGT). *International Journal of Cancer.Journal International Du Cancer.* **132,** 556-567.

Hanke, K., Hohn, O., Bannert, N., 2016. HERV-K(HML-2), a seemingly silent subtenant - but still waters run deep. *APMIS : Acta Pathologica, Microbiologica, Et Immunologica Scandinavica.* **124,** 67-87.

Harlow E. and Lane D. 1988. Antibodies: A laboratory manual. Cold Spring Harbor Laboratory,Cold Spring Harbor, New York

Havecker, E.R., Gao, X., Voytas, D.F., 2004. The diversity of LTR retrotransposons. *Genome Biology.* **5,** 225.

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., Manolio, T.A., 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America.* **106,** 9362-9367.

Hohenadl, C., Germaier, H., Walchner, M., Hagenhofer, M., Herrmann, M., Sturzl, M., Kind, P., Hehlmann, R., Erfle, V., Leib-Mosch, C., 1999. Transcriptional activation of endogenous retroviral sequences in human epidermal keratinocytes by UVB irradiation. *The Journal of Investigative Dermatology.* **113,** 587-594.

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. *Frontiers in Oncology.* **3,** 246.

Hu, T., Zhu, X., Pi, W., Yu, M., Shi, H., Tuan, D., 2017. Hypermethylated LTR retrotransposon exhibits enhancer activity. *Epigenetics.* 0.

Huang, G., Li, Z., Wan, X., Wang, Y., Dong, J., 2013. Human endogenous retroviral K element encodes fusogenic activity in melanoma cells. *Journal of Carcinogenesis.* **12,** 5-3163.109032. Print 2013.

Hurst, T.P. & Magiorkinis, G., 2017. Epigenetic Control of Human Endogenous Retrovirus Expression: Focus on Regulation of Long-Terminal Repeats (LTRs). *Viruses.* **9,** 10.3390/v9060130.

Ilina, T., Labarge, K., Sarafianos, S.G., Ishima, R., Parniak, M.A., 2012. Inhibitors of HIV-1 Reverse Transcriptase-Associated Ribonuclease H Activity. *Biology.* **1,** 521-541.

Jern, P., Sperber, G.O., Ahlsen, G., Blomberg, J., 2005. Sequence variability, gene structure, and expression of full-length human endogenous retrovirus H. *Journal of Virology.* **79,** 6325-6337.

Johanning, G.L., Malouf, G.G., Zheng, X., Esteva, F.J., Weinstein, J.N., Wang-Johanning, F., Su, X., 2017. Expression of human endogenous retrovirus-K is strongly associated with the basal-like breast cancer phenotype. *Scientific Reports.* **7,** 41960.

Kahyo, T., Yamada, H., Tao, H., Kurabe, N., Sugimura, H., 2017. Insertionally polymorphic sites of human endogenous retrovirus-K (HML-2) with long target site duplications. *BMC Genomics.* **18,** 487-017-3872-6.

Kalyanaraman, V.S., Sarngadharan, M.G., Robert-Guroff, M., Miyoshi, I., Golde, D., Gallo, R.C., 1982. A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. *Science (New York, N.Y.).* **218,** 571-573.

Kassiotis, G., 2014. Endogenous retroviruses and the development of cancer. *Journal of Immunology (Baltimore, Md.: 1950).* **192,** 1343-1349.

Katoh, I. & Kurata, S., 2013. Association of endogenous retroviruses and long terminal repeats with human disorders. *Frontiers in Oncology.* **3,** 234.

Kazazian, H.H.,Jr, 2004. Mobile elements: drivers of genome evolution. *Science (New York, N.Y.).* **303,** 1626-1632.

Kim, J., Kollhoff, A., Bergmann, A., Stubbs, L., 2003. Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. *Human Molecular Genetics.* **12,** 233-245.

Kohli, J.S., Tolomio, E., Frigerio, S., Maurichi, A., Rodolfo, M., Bennett, D.C., 2017. Common Delayed Senescence of Melanocytes from Multiple Primary Melanoma Patients. *The Journal of Investigative Dermatology.* **137,** 766-768.

Koito, A. & Ikeda, T., 2013. Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. *Frontiers in Microbiology.* **4,** 28.

Kondo, Y., Shen, L., Cheng, A.S., Ahmed, S., Boumber, Y., Charo, C., Yamochi, T., Urano, T., Furukawa, K., Kwabi-Addo, B., Gold, D.L., Sekido, Y., Huang, T.H., Issa, J.P., 2008. Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. *Nature Genetics.* **40,** 741-750.

Kowalczyk, M.J., Teresiak-Mikolajczak, E., Danczak-Pazdrowska, A., Zaba, R., Adamski, Z., Osmola-Mankowska, A., 2017. Effects of UVA1 Phototherapy on Expression of Human Endogenous Retroviral Sequence (HERV)-K10 gag in Morphea: A Preliminary Study. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research.* **23,** 505-512.

Krishnamurthy, J., Rabinovich, B.A., Mi, T., Switzer, K.C., Olivares, S., Maiti, S.N., Plummer, J.B., Singh, H., Kumaresan, P.R., Huls, H.M., Wang-Johanning, F., Cooper, L.J., 2015. Genetic Engineering of T Cells to Target HERV-K, an Ancient Retrovirus on Melanoma. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research.* **21,** 3241-3251.

Kumaki, Y., Oda, M., Okano, M., 2008. QUMA: quantification tool for methylation analysis. *Nucleic Acids Research.* **36,** W170-5.

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 Consortium, 2001. Initial sequencing and analysis of the human genome. *Nature.* **409,** 860-921.

Landry, J.R., Medstrand, P., Mager, D.L., 2001. Repetitive elements in the 5' untranslated region of a human zinc-finger gene modulate transcription and translation efficiency. *Genomics.* **76,** 110-116.

Lavie, L., Kitova, M., Maldener, E., Meese, E., Mayer, J., 2005. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). *Journal of Virology.* **79,** 876-883.

Lavrentieva, I., Broude, N.E., Lebedev, Y., Gottesman, I.I., Lukyanov, S.A., Smith, C.L., Sverdlov, E.D., 1999. High polymorphism level of genomic sequences flanking insertion sites of human endogenous retroviral long terminal repeats. *FEBS Letters.* **443,** 341-347.

Lee, S.H., Kang, Y.J., Jo, J.O., Ock, M.S., Baek, K.W., Eo, J., Lee, W.J., Choi, Y.H., Kim, W.J., Leem, S.H., Kim, H.S., Cha, H.J., 2014. Elevation of human ERV3-1 env protein expression in colorectal cancer. *Journal of Clinical Pathology.* **67,** 840-844.

Leib-Mösch, C. (2010) Evolutionary history of HERVs in primates (online). Available from: http://www.helmholtz-muenchen.de/en/viro/our-research-projects/retroviral-persistencein-humans/human-endogenous-retroviruses/research-topics/index.html

Lemaitre, C., Harper, F., Pierron, G., Heidmann, T., Dewannieux, M., 2014. The HERV-K human endogenous retrovirus envelope protein antagonizes Tetherin antiviral activity. *Journal of Virology.* **88,** 13626-13637.

Li, Z., Sheng, T., Wan, X., Liu, T., Wu, H., Dong, J., 2010. Expression of HERV-K correlates with status of MEK-ERK and p16INK4A-CDK4 pathways in melanoma cells. *Cancer Investigation.* **28,** 1031-1037.

Liang, S. & Dong, C., 2008. Integrin VLA-4 enhances sialyl-Lewisx/a-negative melanoma adhesion to and extravasation through the endothelium under low flow conditions. *American Journal of Physiology.Cell Physiology.* **295,** C701-7.

Lisanti, S., Omar, W.A., Tomaszewski, B., De Prins, S., Jacobs, G., Koppen, G., Mathers, J.C., Langie, S.A., 2013. Comparison of methods for quantification of global DNA methylation in human cells and tissues. *PloS One.* **8,** e79044.

Liu, C.C., Ou, T.T., Wu, C.C., Li, R.N., Lin, Y.C., Lin, C.H., Tsai, W.C., Liu, H.W., Yen, J.H., 2011. Global DNA methylation, DNMT1, and MBD2 in patients with systemic lupus erythematosus. *Lupus.* **20,** 131-136.

Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., Lorincz, M.C., 2014. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes & Development.* **28,** 2041-2055.

Lower, R., 1999. The pathogenic potential of endogenous retroviruses: facts and fantasies. *Trends in Microbiology.* **7,** 350-356.

Macfarlane, C.M. & Badge, R.M., 2015. Genome-wide amplification of proviral sequences reveals new polymorphic HERV-K(HML-2) proviruses in humans and chimpanzees that are absent from genome assemblies. *Retrovirology.* **12,** 35-015-0162-8.

Manghera, M. & Douville, R.N., 2013. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? *Retrovirology.* **10,** 16-4690-10-16.

Masters, J.R. & Stacey, G.N., 2007. Changing medium and passaging cell lines. *Nature Protocols.* **2,** 2276-2284.

Mastrangelo, G., Krone, B., Fadda, E., Buja, A., Grange, J.M., Rausa, G., de Vries, E., Koelmel, K.F., 2009. Does yellow fever 17D vaccine protect against melanoma? *Vaccine.* **27,** 588-591.

Menendez, L., Benigno, B.B., McDonald, J.F., 2004. L1 and HERV-W retrotransposons are hypomethylated in human ovarian carcinomas. *Molecular Cancer.* **3,** 12.

Moles, J.P., Tesniere, A., Guilhou, J.J., 2005. A new endogenous retroviral sequence is expressed in skin of patients with psoriasis. *The British Journal of Dermatology.* **153,** 83-89.

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,** e70399.

Mourier, T., Mollerup, S., Vinner, L., Hansen, T.A., Kjartansdottir, K.R., Guldberg Froslev, T., Snogdal Boutrup, T., Nielsen, L.P., Willerslev, E., Hansen, A.J., 2015. Characterizing novel endogenous retroviruses from genetic variation inferred from short sequence reads. *Scientific Reports.* **5,** 15644.

Mouse Genome Sequencing Consortium, Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S.E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M.R., Brown, D.G., Brown, S.D., Bult, C., Burton, J., Butler, J., Campbell, R.D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A.T., Church, D.M., Clamp, M., Clee, C., Collins, F.S., Cook, L.L., Copley, R.R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K.D., Deri, J., Dermitzakis, E.T., Dewey, C., Dickens, N.J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D.M., Eddy, S.R., Elnitski, L., Emes, R.D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G.A., Flicek, P., Foley, K., Frankel, W.N., Fulton, L.A., Fulton, R.S., Furey, T.S., Gage, D., Gibbs, R.A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T.A., Green, E.D., Gregory, S., Guigo, R., Guyer, M., Hardison, R.C., Haussler, D., Hayashizaki, Y., Hillier, L.W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D.B., Johnson, L.S., Jones, M., Jones, T.A., Joy, A., Kamal, M., Karlsson, E.K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W.J., Kirby, A., Kolbe, D.L., Korf, I., Kucherlapati, R.S., Kulbokas, E.J., Kulp, D., Landers, T., Leger, J.P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D.R., Mardis, E.R., Matthews, L., Mauceli, E., Mayer, J.H., McCarthy, M., McCombie, W.R., McLaren, S., McLay, K., McPherson, J.D., Meldrim, J., Meredith, B., Mesirov, J.P., Miller, W., Miner, T.L., Mongin, E., Montgomery, K.T., Morgan, M., Mott, R., Mullikin, J.C., Muzny, D.M., Nash, W.E., Nelson, J.O., Nhan, M.N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M.J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K.H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C.S., Poliakov, A., Ponce, T.C., Ponting, C.P., Potter, S., Quail, M., Reymond, A., Roe, B.A., Roskin, K.M., Rubin, E.M., Rust, A.G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M.S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J.B., Slater, G., Smit, A., Smith, D.R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J.P., Von Niederhausern, A.C., Wade, C.M., Wall, M., Weber, R.J., Weiss, R.B., Wendl, M.C., West, A.P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R.K., Winter, E., Worley, K.C., Wyman, D., Yang, S., Yang, S.P., Zdobnov, E.M., Zody, M.C., Lander, E.S., 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature.* **420,** 520-562.

Munoz-Lopez, M. & Garcia-Perez, J.L., 2010. DNA transposons: nature and applications in genomics. *Current Genomics.* **11,** 115-128.

Muster, T., Waltenberger, A., Grassauer, A., Hirschl, S., Caucig, P., Romirer, I., Fodinger, D., Seppele, H., Schanab, O., Magin-Lachmann, C., Lower, R., Jansen, B., Pehamberger, H., Wolff, K., 2003. An endogenous retrovirus derived from human melanoma cells. *Cancer Research.* **63,** 8735-8741.

Nadeau, M.J., Manghera, M., Douville, R.N., 2015. Inside the Envelope: Endogenous Retrovirus-K Env as a Biomarker and Therapeutic Target. *Frontiers in Microbiology.* **6,** 1244.

Nakaoka, T., Saito, Y., Saito, H., 2017. Aberrant DNA Methylation as a Biomarker and a Therapeutic Target of Cholangiocarcinoma. *International Journal of Molecular Sciences.* **18,** 10.3390/ijms18061111.

Nazarian, R., Shi, H., Wang, Q., Kong, X., Koya, R.C., Lee, H., Chen, Z., Lee, M.K., Attar, N., Sazegar, H., Chodon, T., Nelson, S.F., McArthur, G., Sosman, J.A., Ribas, A., Lo, R.S., 2010. Melanomas acquire resistance to B-RAF (V600E) inhibition by RTK or N-RAS upregulation. *Nature.* **468,** 973-977.

Ng, H.H. & Bird, A., 1999. DNA methylation and chromatin modification. *Current Opinion in Genetics & Development.* **9,** 158-163.

Oshlack, A., Robinson, M.D., Young, M.D., 2010. From RNA-seq reads to differential expression results. *Genome Biology.* **11,** 220-2010-11-12-220. Epub 2010 Dec 22.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research.* **29,** e45.

Pergoli, L., Favero, C., Pfeiffer, R.M., Tarantini, L., Calista, D., Cavalleri, T., Angelici, L., Consonni, D., Bertazzi, P.A., Pesatori, A.C., Landi, M.T., Bollati, V., 2014. Blood DNA methylation, nevi number, and the risk of melanoma. *Melanoma Research.* **24,** 480-487.

Rastogi, A., Ali, A., Tan, S.H., Banerjee, S., Chen, Y., Cullen, J., Xavier, C.P., Mohamed, A.A., Ravindranath, L., Srivastav, J., Young, D., Sesterhenn, I.A., Kagan, J., Srivastava, S., McLeod, D.G., Rosner, I.L., Petrovics, G., Dobi, A., Srivastava, S., Srinivasan, A., 2016. Autoantibodies against oncogenic ERG protein in prostate cancer: potential use in diagnosis and prognosis in a panel with C-MYC, AMACR and HERV-K Gag. *Genes & Cancer.* **7,** 394-413.

Reid, Y., Storts, D., Riss, T. and Minor, L., 2004. Authentication of Human Cell Lines by STR DNA Profiling Analysis. In Sittampalam, G. S., Coussens, N. P., Brimacombe, K., Grossman, A., Arkin, M., Auld, D., Austin, C., Bejcek, B., Glicksman, M., Inglese, J., Iversen, P. W., Li, Z., McGee, J., McManus, O., Minor, L., Napper, A., Peltier, J. M., Riss, T., Trask OJ, J. and Weidner, J., eds, *Assay Guidance Manual.* Bethesda (MD).

Reis, B.S., Jungbluth, A.A., Frosina, D., Holz, M., Ritter, E., Nakayama, E., Ishida, T., Obata, Y., Carver, B., Scher, H., Scardino, P.T., Slovin, S., Subudhi, S.K., Reuter, V.E., Savage, C., Allison, J.P., Melamed, J., Jager, E., Ritter, G., Old, L.J., Gnjatic, S., 2013. Prostate cancer progression correlates with increased humoral immune response to a human endogenous retrovirus GAG protein. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research.* **19,** 6112-6125.

Rhyu, D.W., Kang, Y.J., Ock, M.S., Eo, J.W., Choi, Y.H., Kim, W.J., Leem, S.H., Yi, J.M., Kim, H.S., Cha, H.J., 2014. Expression of human endogenous retrovirus env genes in the blood of breast cancer patients. *International Journal of Molecular Sciences.* **15,** 9173-9183.

Riverso, M., Montagnani, V., Stecca, B., 2017. KLF4 is regulated by RAS/RAF/MEK/ERK signaling through E2F1 and promotes melanoma cell growth. *Oncogene.*

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. *Nature Biotechnology.* **29,** 24-26.

Romanish, M.T., Cohen, C.J., Mager, D.L., 2010. Potential mechanisms of endogenous retroviral-mediated genomic instability in human cancer. *Seminars in Cancer Biology.* **20,** 246-253.

Roulois, D., Loo Yau, H., Singhania, R., Wang, Y., Danesh, A., Shen, S.Y., Han, H., Liang, G., Jones, P.A., Pugh, T.J., O'Brien, C., De Carvalho, D.D., 2015. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell.* **162,** 961-973.

Sadakierska-Chudy, A. & Filip, M., 2015. A comprehensive view of the epigenetic landscape. Part II: Histone post-translational modification, nucleosome level, and chromatin regulation by ncRNAs. *Neurotoxicity Research.* **27,** 172-197.

Saito, Y., Nakaoka, T., Saito, H., 2017. Chapter Eight - A New Molecular Mechanism Underlying the Antitumor Effect of DNA Methylation Inhibitors via an Antiviral Immune Response. *Advances in Protein Chemistry and Structural Biology.* **106,** 227-242.

Saito, Y., Suzuki, H., Tsugawa, H., Nakagawa, I., Matsuzaki, J., Kanai, Y., Hibi, T., 2009. Chromatin remodeling at Alu repeats by epigenetic treatment activates silenced microRNA-512-5p with downregulation of Mcl-1 in human gastric cancer cells. *Oncogene.* **28,** 2738-2744.

Schmitt, K., Reichrath, J., Roesch, A., Meese, E., Mayer, J., 2013. Transcriptional profiling of human endogenous retrovirus group HERV-K(HML-2) loci in melanoma. *Genome Biology and Evolution.* **5,** 307-328.

Schug, Z., Peck, B., Jones, D., Zhang, Q., Grosskurth, S., Alam, I., Goodwin, L., Smethurst, E., Mason, S., Blyth, K., McGarry, L., James, D., Shanks, E., Kalna, G., Saunders, R., Jiang, M., Howell, M., Lassailly, F., Thin, M., Spencer-Dene, B., Stamp, G., van den Broek, N.F., Mackay, G., Bulusu, V., Kamphorst, J., Tardito, S., Strachan, D., Harris, A., Aboagye, E., Critchlow, S., Wakelam, M.O., Schulze, A., Gottlieb, E., 2015. Acetyl-CoA Synthetase 2 Promotes Acetate Utilization and Maintains Cancer Cell Growth under Metabolic Stress. *Cancer Cell.* **27,** 57-71.

Serafino, A., Balestrieri, E., Pierimarchi, P., Matteucci, C., Moroni, G., Oricchio, E., Rasi, G., Mastino, A., Spadafora, C., Garaci, E., Vallebona, P.S., 2009a. The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation. *Experimental Cell Research.* **315,** 849-862.

Sheen, F.M., Sherry, S.T., Risch, G.M., Robichaux, M., Nasidze, I., Stoneking, M., Batzer, M.A., Swergold, G.D., 2000. Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. *Genome Research.* **10,** 1496-1508.

Sicat, J., Sutkowski, N., Huber, B.T., 2005. Expression of human endogenous retrovirus HERV-K18 superantigen is elevated in juvenile rheumatoid arthritis. *The Journal of Rheumatology.* **32,** 1821-1831.

Singh, S., Kaye, S., Gore, M.E., McClure, M.O., Bunker, C.B., 2009. The role of human endogenous retroviruses in melanoma. *The British Journal of Dermatology.* **161,** 1225-1231.

Stengel, S., Fiebig, U., Kurth, R., Denner, J., 2010. Regulation of human endogenous retrovirus-K expression in melanomas by CpG methylation. *Genes, Chromosomes & Cancer.* **49,** 401-411.

Strahl, B.D. & Allis, C.D., 2000. The language of covalent histone modifications. *Nature.* **403,** 41-45.

Strissel, P.L., Ruebner, M., Thiel, F., Wachter, D., Ekici, A.B., Wolf, F., Thieme, F., Ruprecht, K., Beckmann, M.W., Strick, R., 2012. Reactivation of codogenic endogenous retroviral (ERV) envelope genes in human endometrial carcinoma and prestages: Emergence of new molecular targets. *Oncotarget.* **3,** 1204-1219.

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,** 90-4690-8-90.

Sukapan, P., Promnarate, P., Avihingsanon, Y., Mutirangura, A., Hirankarn, N., 2014. Types of DNA methylation status of the interspersed repetitive sequences for LINE-1, Alu, HERV-E and HERV-K in the neutrophils from systemic lupus erythematosus patients and healthy controls. *Journal of Human Genetics.* **59,** 178-188.

Suntsova, M., Garazha, A., Ivanova, A., Kaminsky, D., Zhavoronkov, A., Buzdin, A., 2015. Molecular functions of human endogenous retroviruses in health and disease. *Cellular and Molecular Life Sciences: CMLS.* **72,** 3653-3675.

Thompson, P.J., Macfarlan, T.S., Lorincz, M.C., 2016. Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Molecular Cell.* **62,** 766-776.

Timar, J., Vizkeleti, L., Doma, V., Barbai, T., Raso, E., 2016. Genetic progression of malignant melanoma. *Cancer Metastasis Reviews.* **35,** 93-107.

Tomilin, N.V., 1999. Control of genes by mammalian retroposons. *International Review of Cytology.* **186,** 1-48.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., Issa, J.P., 1999. CpG island methylator phenotype in colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America.* **96,** 8681-8686.

Tran, K.A., Cheng, M.Y., Mitra, A., Ogawa, H., Shi, V.Y., Olney, L.P., Kloxin, A.M., Maverakis, E., 2015. MEK inhibitors and their potential in the treatment of advanced melanoma: the advantages of combination therapy. *Drug Design, Development and Therapy.* **10,** 43-52.

Urnovitz, H.B. & Murphy, W.H., 1996. Human endogenous retroviruses: nature, occurrence, and clinical implications in human disease. *Clinical Microbiology Reviews.* **9,** 72-99.

Wallace, T.A., Downey, R.F., Seufert, C.J., Schetter, A., Dorsey, T.H., Johnson, C.A., Goldman, R., Loffredo, C.A., Yan, P., Sullivan, F.J., Giles, F.J., Wang-Johanning, F., Ambs, S., Glynn, S.A., 2014. Elevated HERV-K mRNA expression in PBMC is associated with a prostate cancer diagnosis particularly in older men and smokers. *Carcinogenesis.* **35,** 2074-2083.

Wang-Johanning, F., Li, M., Esteva, F.J., Hess, K.R., Yin, B., Rycaj, K., Plummer, J.B., Garza, J.G., Ambs, S., Johanning, G.L., 2014. Human endogenous retrovirus type K antibodies and mRNA as serum biomarkers of early-stage breast cancer. *International Journal of Cancer.* **134,** 587-595.

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. *International Journal of Cancer.Journal International Du Cancer.* **120,** 81-90.

Weiss, R.A., 2013. On the concept and elucidation of endogenous retroviruses. *Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences.* **368,** 20120494.

Weiss, R.A., 2006. The discovery of endogenous retroviruses. *Retrovirology.* **3,** 67-4690-3-67.

Wildschutte, J.H., Williams, Z.H., Montesion, M., Subramanian, R.P., Kidd, J.M., Coffin, J.M., 2016. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proceedings of the National Academy of Sciences of the United States of America.* **113,** E2326-34.

Willmore-Payne, C., Holden, J.A., Tripp, S., Layfield, L.J., 2005. Human malignant melanoma: detection of BRAF- and c-kit–activating mutations by high-resolution amplicon melting analysis. *Human Pathology.* **36,** 486-493.

Wolffe, A.P. & Matzke, M.A., 1999. Epigenetics: regulation through repression. *Science (New York, N.Y.).* **286,** 481-486.

Wylie, A., Jones, A.E., D'Brot, A., Lu, W.J., Kurtz, P., Moran, J.V., Rakheja, D., Chen, K.S., Hammer, R.E., Comerford, S.A., Amatruda, J.F., Abrams, J.M., 2016. P53 Genes Function to Restrain Mobile Elements. *Genes & Development.* **30,** 64-77.

Xu, L., Shen, S.S., Hoshida, Y., Subramanian, A., Ross, K., Brunet, J.P., Wagner, S.N., Ramaswamy, S., Mesirov, J.P., Hynes, R.O., 2008. Gene expression changes in an animal melanoma model correlate with aggressiveness of human melanoma metastases. *Molecular Cancer Research: MCR.* **6,** 760-769.

Yoder, J.A., Walsh, C.P., Bestor, T.H., 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics: TIG.* **13,** 335-340.

Yoo, C.B. & Jones, P.A., 2006. Epigenetic therapy of cancer: past, present and future. *Nature Reviews.Drug Discovery.* **5,** 37-50.

Young, G.R., Stoye, J.P., Kassiotis, G., 2013. Are human endogenous retroviruses pathogenic? An approach to testing the hypothesis. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology.* **35,** 794-803.

Zhang, Y. & Reinberg, D., 2001. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes & Development.* **15,** 2343-2360.

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.* **7,** 84093-84117.