The cellular splicing regulator SRSF2 controls HPV16 E6 mRNA stability and contributes to the cervical tumour phenotype.

By

Melanie McFarlane

A thesis presented for the degree of Doctor of Philosophy

In

College of Medical Veterinary and Life Sciences

At the University of Glasgow

September 2011

MRC-University of Glasgow

Centre for Virus Research

Institute of Infection Immunity and Inflammation

College of Medical Veterinary and Life Sciences

University of Glasgow

Acknowledgements

Firstly I would like to thank my supervisor Dr Sheila Graham for all her invaluable support, understanding, guidance and advice. I would never have managed to complete this Ph.D. without her constant support. I would also like to thank her for critically reading this thesis. A huge thank you also goes to all the members of the Graham group who have helped and encouraged me, especially Al and Tanya to whom I will always be grateful for their help and humour. I would also like to thank Prof Iain Morgan for supervising my placement during my first year rotation and for his continuing interest in my studies. I would particularly like to thank my assessors Dr Patricia Martin and Prof Jeremy Mottram for their input and suggestions.

Finally my thanks go to my family for their never ending love and support. To my Mum, Dad and Stephanie who have had to endure the ups and downs with me especially during the first and final years and have never wavered in their support. To my Granparents who have encouraged me and provided me with free meals, thank you! And to Len, without whom this thesis would never have been completed, thank you for all your love and encouragement over the last four years. This thesis is as much yours as it is mine.

The author received funding from the MRC Doctoral Training Scheme and unless otherwise stated all the work included in this thesis was carried out by the author.

Abstract

HPV16 is a double stranded DNA virus which has a close association with cervical cancer development. HPV16 expresses two oncoproteins E6 and E7, which when overexpressed transform the virus-infected epithelial cells. E6 and E7-encoding RNAs have been shown to be alternatively spliced to give at least four mRNA isoforms (E6 full length, E6*I, E6*II, E6*X) however it is not known what functions any putative alternative oncoproteins or indeed, RNA isoforms may have. Alternative splicing is regulated by host cell factors, namely SR proteins and hnRNP proteins. SR proteins are the positive splicing regulators in the cell and generally promote splicing of both cellular and viral RNAs. It is not known which RNA splicing factors are required for E6/E7 RNA processing during infection and transformation. HPV16 viral RNAs are extensively alternatively spliced therefore identification of the host factors involved in the processing of viral RNAs could have therapeutic benefits because completion of the virus life cycle requires alternative splicing and if this could be prevented by targeting of the splicing factors involved, then the virus infection could be prevented. Similarly if splicing is altered upon transformation of the cervical epithelial cells, then prevention of this alteration in splicing could perhaps inhibit virus-induced transformation of the cervical epithelial cells. Recently it has been shown that overexpression of SR proteins can lead to cellular transformation and certain SR proteins have already been reported to be upregulated in some cancers. Therefore the focus of this PhD was to investigate the expression of the alternative E6/E7 RNA isoforms and identify the protein (s) responsible for their processing.

The first aim was to investigate the expression of the E6/E7 RNA isoforms in virus-infected cells and during cervical epithelial cell transformation and to try to assign any function to the individual isoforms. E6 and E7 are encoded by the viral early genes but are expressed in both undifferentiated and differentiated virus infected cells. Therefore I investigated whether the expression of the isoforms was altered during differentiation. RT-PCR experiments were carried out using RNA extracted from undifferentiated and differentiated W12E cells (HPV16-infected cervical epithelial cells). Results indicated that E6/E7 splicing is not altered upon epithelial differentiation however E6 and E7 mRNA abundance increased upon differentiation. Furthermore, experiments carried out in W12G cells, where HPV16 genomes are integrated into the host genome and no viral life cycle is taking place, suggested that the increase in E6 and E7 RNA expression was not due to cell differentiation, but due to a virus-induced increase in expression. Further RT-PCR experiments in HPV16 transformed cells lines demonstrated that E6/E7 RNA isoform expression is altered as cervical epithelial cells become tumourigenic. Small E6 isoforms, E6*II and E6*X are upregulated in virus transformed cells suggesting a tumour promoting function for the isoforms. To test this, E6/E7 isoform-expressing constructs were created and transiently transfected into HPV-negative C33a cervical cancer cells and the proliferation rate and ability to form colonies in soft agar investigated. Compared to the longer E6/E7 isoforms the two smallest isoforms promoted cellular proliferation as the cell growth rate increased. However anchorage independent growth assays were inconclusive suggesting there may be a combinatorial effect of the E6/E7 isoforms on transformation of the cells.

My next aim was to investigate the expression of SR proteins during transformation of HPV16-infected epithelial cells. Western blot and

immunohistochemical analysis showed that SR proteins SRSF1-3 were specifically upregulated upon cervical epithelial transformation. For SRSF2 and 3, this was not due to gene amplification as qPCR analysis of gene copy number showed no significant difference in CT values between the W12 cell lines suggesting that upregulation of SRSF2 and 3 may be at a transcriptional level. However there was a significant difference in SRSF1 gene copy number that may account for its upregulation.

My final aim was to identify the SR protein (s) responsible for the alteration in E6/E7 isoform expression in HPV16-transformed cells. This was achieved by siRNA depletion of the overexpressed SR proteins and RT-PCR of E6/E7 RNA. Surprisingly, none of the SR protein knockdowns resulted in any detectable alteration of RNA isoforms. However, SRSF2 knockdown specifically resulted in a significant reduction in all E6/E7 encoding RNAs. Moreover, after SRSF2 knockdown, p53 levels were increased suggesting an impairment of E6 protein function. The reduction in E6/E7 RNA was not due to a decrease in transcription as demonstrated by transcription assays utilising an HPV16 LCR luciferase reporter. Interestingly, E6/E7 RNA stability assays showed that RNA half life is reduced when SRSF2 is knocked down. SRSF1 and 2 have previously been shown to be oncogenic in breast and ovarian cancers respectively. So the effects of SRSF2 knockdown on cell growth rate, colony formation, apoptosis entry and cell cycle were analysed in transformed cervical epithelial cells. The results from these experiments indicated that overexpression of SRSF2 in cervical epithelial cells is tumour promoting. My data clearly indicates that SRSF2 should be considered to be a proto-oncogene.

Table of contents

List of figures

List of tables

Abbreviations

APS	Ammonium persulphate
bp	base pair
BPV	bovine papillomavirus
CIN	cervical intraepithelial neoplasia
CMV	cytomegalovirus
CPSF	cleavage and polyadenylation specificity factor
Cst-F	cleavage stimulatory factor
CTD	carboxy terminal domain
DAPI	4', 6'-diamino-2-phenylindole hydrochloride
DCS	donor calf serum
DEPC	diethyl pyrocarbonate
DMEM	Dulbecco's modified eagle medium
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
ECL	enhanced chemiluminesence
EDTA	ethylenediaminetre-acetic acid
EGF	epidermal growth factor
ESE	exonic splicing enhancer
ESS	exonic splicing suppressor
FCS	foetal calf serum
GFP	green fluorescent protein

HEPES	N-2-hydroxyethyl piperazine-N'-2-ethane sulphonic acid
hnRNP	heterogeneous nuclear ribonucleoprotein
HPV	human papillomavirus
IDC	indole derivative compound
LCR	long control region
mRNA	messenger RNA
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PBS-T	PBS-Tween
PCR	polymerase chain reaction
PV	papillomavirus
RT-PCR	reverse transcriptase PCR
qRT-PCR	quantitative RT-PCR
RNA	ribonucleic acid
rpm	revolutions per minute
RT	reverse transcriptase
SDS	sodium dodecyl sulphate
snRNP	small nuclear ribonucleoprotein
SR	serine and arginine rich
SV40	simian virus 40
TBE	Tris-borate EDTA buffer
TE	Tris-EDTA buffer

µ	micro
U	units
U2AF	U2 auxiliary factor
UTR	untranslated region
v/v	volume per volume
w/v	weight per volume

Introduction

1 Papillomaviruses

1.1 Classification

Papillomaviruses (PVs) are small double stranded DNA viruses with genomes of around 8kb in size. PVs infect mucosal and cutaneous epithelial cells and cause benign lesions such as warts and veruccas. Human papillomaviruses (HPVs) have also been linked to some malignancies including cervical and head and neck cancers. Papillomaviruses are classified according to homology between their L1 gene sequences, as the L1 gene has been found to be the most conserved region of the genome (de Villiers et al., 2004). A new virus is identified if the L1 region is more than ten percent different from the most similar viral sequence. If the coding sequence is between two and ten percent different, the unidentified virus is considered to be a new subtype of the most similar virus and if the difference observed is less than two percent the new virus is considered a variant (Bernard, 2005). Papillomaviruses can infect a very wide range of host species including birds, reptiles and mammals including humans, cattle and rabbits.

There have been over 120 human types identified which can be divided up into groups of virus types called genera (de Villiers et al., 2004). PVs can be divided into sixteen genera of viruses identified by Greek letters e.g. alpha-papillomaviruses (Bernard, 2005). Genera are composed of viruses that are involved in similar types of lesions e.g. alpha papillomaviruses all cause mucosal or genital lesions and are clinically the most important PV genus (de Villiers et

al., 2004). Each genus can then be further divided up into species of PV, where closely related viruses can be grouped together, as very often closely related viruses have similar pathological features (Bernard, 2005). The largest genus by far is the alpha-papillomavirus genus which contains 15 species of PV and 60 different types. Alpha papillomaviruses that contain the most common human infective viruses can then be further divided into two groups termed “high-risk” and “low-risk” viruses depending on their ability to lead to malignancies. HPV16, HPV18 and others that cause cervical cancer are members of the “high-risk” alpha-PV genus.

The second largest genus is the beta-PV genus, containing 5 species with 25 different types. Beta-PVs are cutaneous-specific viruses and are therefore not associated with cervical cancer. Beta-papillomaviruses are found mainly in normal skin. They can however be found associated with skin tumours particularly in immunosuppressed or immunocompromised people (Harwood et al., 2000). Some beta-PVs are associated with Epidermodysplasia Verruciformis (EV), a rare autosomal recessive genetic condition that results in flat papillomavirus associated lesions on sun-exposed skin of young people and can lead to squamous cell carcinoma (SCC) (Sterling, 2005). HPV5 and HPV8 are commonly found in EV related SCCs and are examples of beta-PV genus members (Sterling, 2005). Gamma-PV, Mu-PV and Nu-PV genera also contain HPVs, however these viruses are less well defined than the alpha and beta-PVs. The remaining 11 genera contain PVs from various mammals, birds and reptiles. The papillomavirus phylogenetic tree is shown in Figure 1.1.

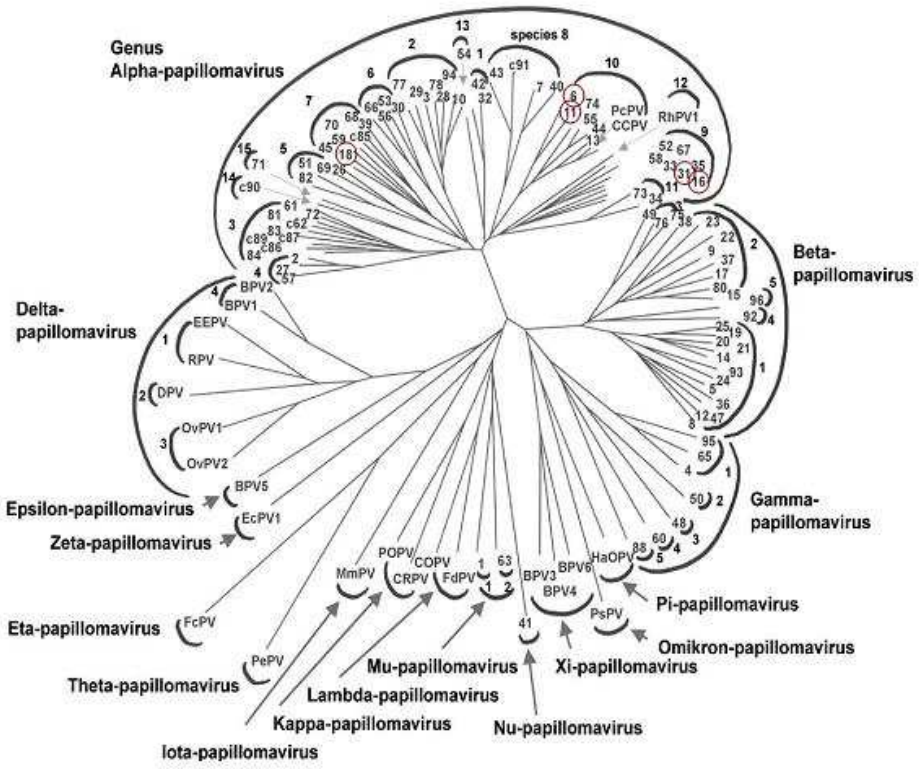


Figure 1.1 Phylogenetic tree of the papillomavirus genus. Adapted from (de Villiers et al., 2004). The most common anogenital infective HPVs, HPV16, HPV18, HPV31, HPV11 and HPV6 are indicated by red circles.

1.2 Pathogenicity

HPVs target the epithelial tissue at three main sites, the anogenital tract, the respiratory tract and the skin. After initial infection, the virus genome moves to the cell nucleus where it forms an episome and is stably maintained at around 50 – 100 copies per cell (Stubenrauch & Laimins, 1999). This state is considered to be a persistent infection. During the productive infection where viral DNA is episomally maintained, infectious virions are produced. This productive infection usually results in the formation of benign, warty lesions. These benign lesions normally regress spontaneously without treatment, due to the presence of circulating $CD4^+$ and $CD8^+$ T lymphocytes (Stanley, 2001). Antibodies to the L1 capsid protein can be found in people with previous HPV infections these antibodies are found to provide long term protection and are the basis for the recently licensed HPV vaccine (Stanley, 2001). However, during infection with some HPVs the circular genome can break and integrate into the host cell chromosome due to DNA damage. DNA damage probably occurs after encounters with carcinogens in the case of mucosal epithelia and UV in the case of cutaneous epithelia. HPVs with the capacity to integrate are termed high-risk HPVs. This integration of the viral genome can result in malignancy. High-risk HPVs are viruses that when they infect epithelial cells have the capacity to integrate their genome into the host chromosome and therefore result in uncontrolled cell proliferation, leading to tumour formation. There is increasing evidence that integration does not always need to occur to lead to malignancy (Gray et al., 2010). Low-risk HPVs do not have this capacity and therefore infection with low-risk viruses rarely leads to cancer development. HPV16 and

18 are the most well known examples of high-risk viruses and HPV2 and 6 are examples of low-risk HPVs. Beta-PVs are not normally further divided into low or high risk PVs, however some can cause cancer without genome integration. This happens by different mechanisms than alpha-PVs, which will not be discussed here.

1.3 Vaccine

There has long been a great interest in developing a vaccination to protect against HPV infection. Original studies identified the protective ability of virus like particles (VLP) which are composed of L1 and L2 proteins, but without any viral DNA, making them non-infective. L1 and L2 are highly immunogenic proteins which the virus only expresses away from the immune surveillance processes which do not occur in the granular layer of the epithelium. The first demonstration of the protection provided by these VLPs was done using cotton-tail rabbit papillomavirus, where rabbits vaccinated with L1 alone or L1-L2 VLPs developed fewer papillomas than control animals (Breitburd et al., 1995). The first study using mucosal infective papillomaviruses were carried out using BPV4 VLPs in cattle where again, the vaccinated animals developed fewer papillomas than control animals (Kirnbauer et al., 1996). Recently there has been a vaccine licensed for use to prevent HPV infection. The vaccine licensed in the UK is called Cervarix and contains HPV16 and HPV18 VLPs which elicit an immune response against L1 only.

While this is obviously a huge leap in prevention of HPV16 and 18 infections and cervical cancers there is still a long way to go before HPV infections are

eliminated. While the vaccine has been approved for use, there is no data on the effectiveness of the vaccine in long term protection from infection. While the UK immunises with a bivalent vaccine there is also a quadravalent vaccine licensed which protects not only against HPV16 and 18 infections but also HPV6 and 11. The choice of bivalent vaccine in the UK is largely an economic decision which may in hindsight prove to be a mistake. The UK is only vaccinating girls, which seems strange as boys can not only become infected with the viruses but penile, anal and increasingly oral carcinomas have been linked to high-risk HPV infection. Finally as the vaccine elicits an immune response against the L1 protein, only the initial infection is prevented and the vaccine will have no effect for those people already infected with HPV. There are some concerns that while vaccination against the most common high-risk viruses may eliminate these viruses, there are other HPVs that infect the anogenital epithelium and have the potential to lead to malignancies and that these viruses may fill the niche left by elimination of HPV16 and 18. HPV16 and HPV18 are the most common causes of cervical cancers and if they are eliminated other high-risk viruses such as HPV31, 51 or 55 may take over and cause cancers. It may be that HPV16 and 18 are the most common cause of cervical cancer because they are the fittest viruses and can out-compete the other high-risk viruses where there is an infection with multiple virus types. With HPV16 and 18 eliminated the other high risk viruses may cause many more cervical cancers than they currently do.

1.4 HPV Life Cycle and Epithelial Differentiation

PVs must infect the basal epithelial layer in order to initiate infection (Doorbar, 2005). This requires the epithelium to become damaged resulting in a disruption of the epithelial barrier allowing the virus access to the lower basal layer. It has been suggested that infection of stem cells within the epithelium may be required for lesion persistence (Schmitt et al., 1996). Recently it has been demonstrated, using a rabbit oral papillomavirus (ROPV) infection as a model system for low risk HPV infection, that low levels of viral DNA and RNA can be detected in previously infected basal stem cells up to one year after the initial infection has regressed, however no late protein expression was detected indicating that no productive infection was taking place (Maglennon et al., 2011). Viral DNA and RNA could not be detected in the surrounding normal epithelium indicating the virus may set up latency in particular basal stem cells (Maglennon et al., 2011). This pool of latent ROPV DNA could be evidence of persistence and perhaps the virus could then be stimulated to reactivate by external factors (Maglennon et al., 2011). Stem cells can be found within hair follicles and replicate in the interfollicular epithelium, therefore for cutaneous specific PVs e.g. HPV5, this may be an important location of infection (Boxman et al., 2001, Doorbar, 2005). Similarly, Schmitt et al. (1996) demonstrated that Cotton-tail Rabbit Papillomavirus (CRPV) infect and persist in epidermal skin cells with stem cell like properties within the hair follicle. There is still some controversy about the viral entry receptor and this will be discussed later. After initial receptor binding the virus is then thought to enter the cell via a clathrin-mediated endocytosis process (Selinka et al., 2002). As the basal cells start to differentiate and move up through the skin layers, the PV life cycle begins.

1.4.1 Normal Epithelial Cell Cycle

In normal epithelial tissues the basal cell layer is the only layer where DNA synthesis occurs. In this layer proliferation associated keratins such as K5 and K14 are expressed. After cell division one daughter cell remains in the basal layer while the other leaves the basal layer and enters the spinous layer. In this layer, the transit amplifying cells exit the cell cycle and begin to differentiate, and while in the spinous layer and the cells begin to accumulate keratins (Stubenrauch & Laimins, 1999). This layer is mitotically inactive and the major role of the spinous layer cells is to produce more keratins (Fuchs, 1990). As the cells differentiate the keratin expression profile is altered. Proliferation associated keratins are switched off and differentiation associated keratins such as K10 and K1 are expressed. Finally, as the cells move up through the spinous layer and enter the granular layer they begin to produce envelope proteins such as Involucrin and Filaggrin (Fuchs, 1990). After reaching the granular layer the cells stop producing keratins. Once the cornified layer is reached the cells then loose the cell nucleus, essentially becoming dead keratin-containing cells (Stubenrauch & Laimins, 1999). These cells are then sloughed off naturally allowing a continuous cycle of replacement from the basal layer. The whole epithelium differentiation program takes around 14 days to complete. The structure of the normal epithelium is shown in Figure 1.2.

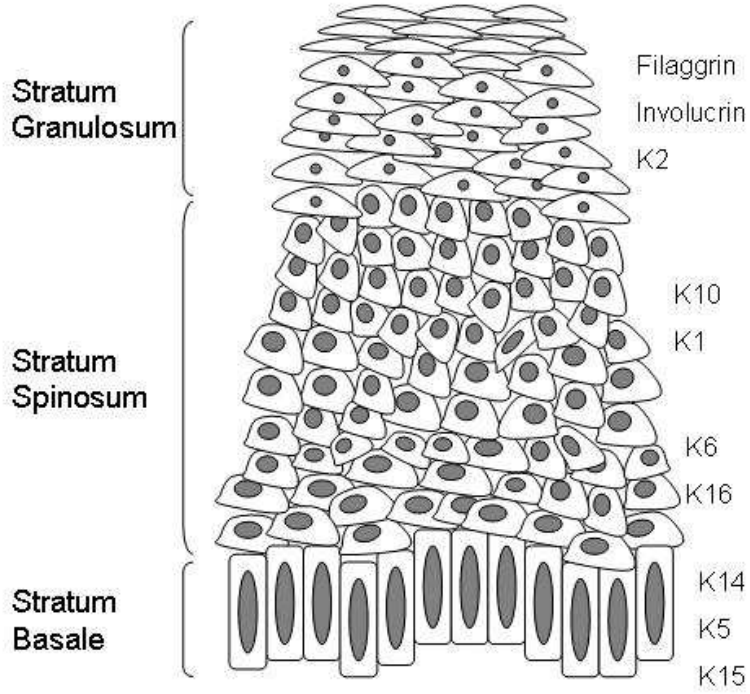


Figure 1.2 Schematic diagram of an uninfected cervical epithelium. Actively dividing cells are located in the stratum basale. As transit amplifying cells move up through the epithelium into the stratum spinosum they stop dividing and undergo differentiation. Terminally differentiated cells are located in the stratum granulosum and are naturally sloughed off. “K”, Keratin.

1.4.2 PV Infected Epithelial Cell Cycle

After infection of basal epithelial cells the PV genome is deposited in the nucleus and maintained as viral episomes. Maintenance of the viral episome requires the early viral proteins E1 and E2. E1 is the protein responsible for initiating the viral DNA replication and E2 is involved with regulating the transcription of viral RNAs among other functions (Phelps & Howley, 1987). E2 is also the viral protein which can bind to mitotic chromosomes via a tethering protein, therefore allowing the distribution of the viral genomes along with the normal separation of chromosomes between dividing cells (You et al., 2004, Parish et al., 2006a, Donaldson et al., 2007). During the initial stages of infection only the viral early genes are expressed. The other viral early genes E6 and E7 are responsible for driving cellular proliferation and avoiding apoptosis (Dyson et al., 1989, Crook et al., 1991). As the infected cell moves up through the cell layers there is some evidence that it may become locked in the G2 stage of the cell cycle by action of E1^E4 (Davy et al., 2002, Davy et al., 2005), preventing differentiation and allowing amplification of viral DNA to continue before packaging of the genomes into the newly synthesised viral capsids occurs (Doorbar, 2005). However there is also evidence that mitosis occurs within lesions, and in fact this is one major clinical sign of a PV induced lesion. Therefore this is still a grey area of papillomavirus research. Blockage of cell cycle exit, by E1^E4, may allow cellular replicative enzymes to remain active and allow viral replication (Doorbar, 2005). During the amplification stage of the PV life cycle other promoters that are late in infection (Milligan et al., 2007), found throughout the viral genome, may become active, increasing the

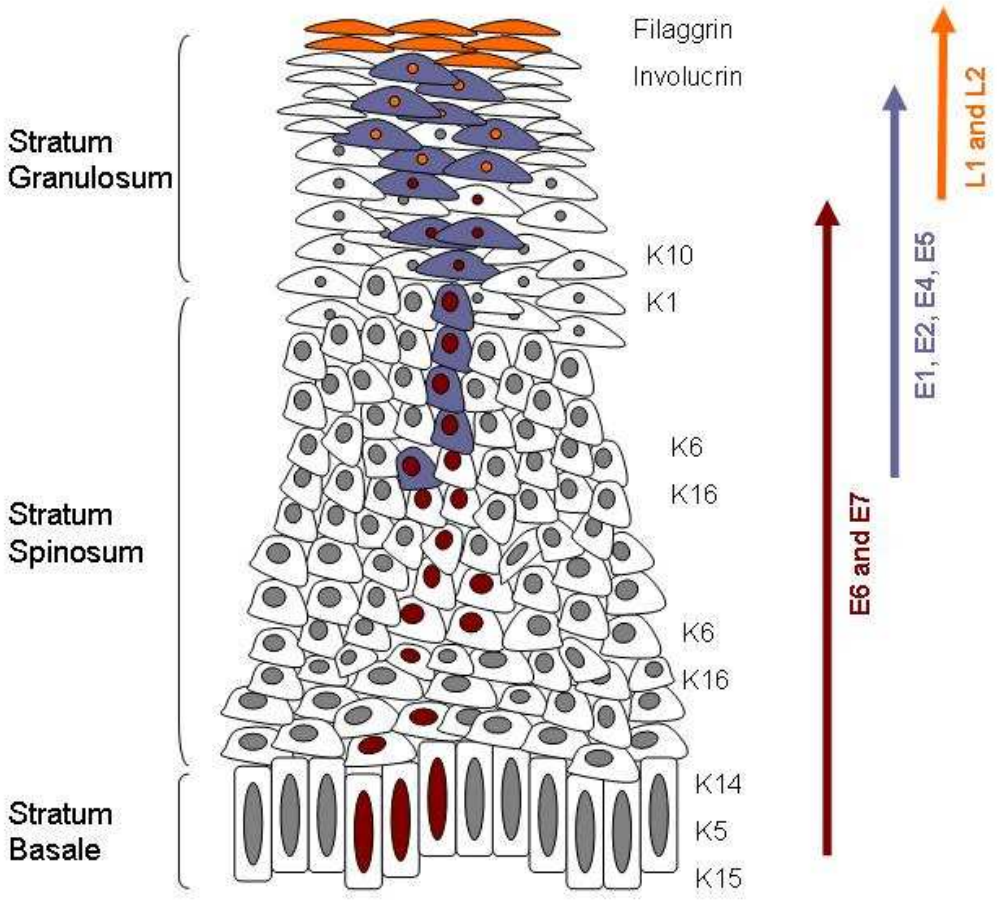


Figure 1.3 Schematic diagram of an HPV-infected cervical epithelium. HPVs enter the basal layer through wounds in the epithelium and the virus infects actively dividing cells in the stratum basale. Only viral early proteins E6 and E7 (red) and E1 and E2 are expressed in the basal layer. As the cells move up through the epithelium into the stratum spinosum they are stimulated by the virus to continue dividing where no division should be taking place resulting in the characteristic "wart". Differentiation is postponed in this layer to allow the virus to amplify its genetic material. All early genes (red and blue) are expressed in this layer. Virus late structural proteins (yellow) are only expressed in terminally differentiated cells in the stratum granulosum and the cells are naturally sloughed off, releasing the newly formed virus particles. "K", Keratin.

levels of transcription of E4 and E5 (Middleton et al., 2003, Doorbar, 2005) and eventually the other late genes, L1 and L2 which encode for the viral structural proteins (Doorbar & Gallimore, 1987). E5 is responsible for blocking the immune system activation in response to HPV infection by preventing the MHC class I and II molecules from presenting to the immune system (Zhang et al., 2003, Ashrafi et al., 2005). The late PV proteins are only expressed once the infected cell reaches the granular layer because these proteins are immunogenic, and form the basis of the HPV vaccine; therefore they must be hidden from the immunosurveillence processes in the epithelium. L1 proteins can induce a neutralising T cell response which is why they are hidden from the immune system (reviewed by (Stanley, 2008). The newly formed viral particles are released as the cells reach the stratum corneum and are naturally sloughed off. The release of the viral particles by sloughing of the epithelial cells provides protection for the virus from the environment. The epithelium acts as a barrier to infection and protects from the external environment. Infection of the epithelium by HPV interferes with these epithelial functions as the tightly regulated epithelial structure is disrupted by wart formation. The structure of the papillomavirus infected epithelium is shown in Figure 1.3.

1.5 Gene Expression

HPV gene expression is tightly regulated and linked to epithelial differentiation. The expression of viral genes can be divided into two stages, early gene expression and late gene expression. Early genes are gene that are expressed as soon as the virus enters the basal epithelial cells and establishes its genome as an episome. Late proteins are expressed only as the infected cell becomes differentiated as these proteins are often highly immunogenic and can only be expressed away from the immune surveillance. The viral genes are named according to where in the viral life cycle they are expressed. There are eight viral genes, E1, E2, E4, E5, E6, E7, L1 and L2 where E stands for early and L stands for late genes. HPV16 has two main viral promoters; P_{97} which is activated early in infection and from which the majority of the early genes are expressed and P_{670} which is activated upon differentiation and from which late genes are expressed. HPV16 also has three polyadenylation sites, one just after the E5 ORF at nucleotide 4215 which is used during early gene expression and two just after the L1 ORF at nucleotide 7321 and are used during late gene expression. There are also other promoters that can be found throughout the viral genome (Milligan et al., 2007). The linear HPV16 genome is shown in Figure 1.4.

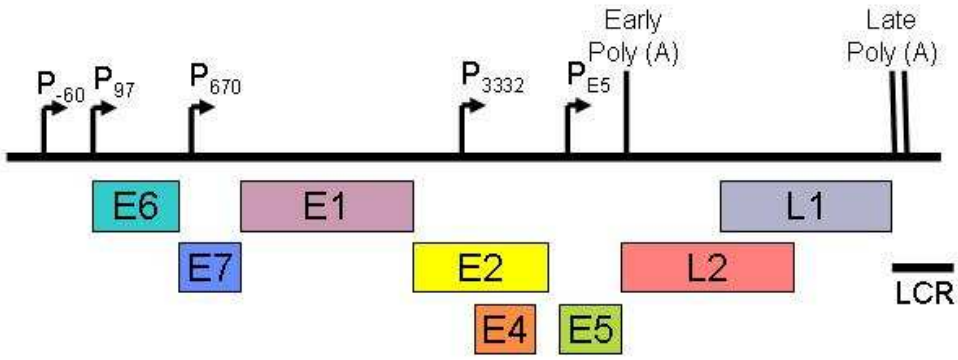


Figure 1.4 Schematic diagram of the linear HPV16 genome. The major early (P97) and late promoter (P670) are indicated by arrows along with other promoters identified by Milligan et al. (2007). The early and the two late polyadenylation sites are indicated by vertical bars. The position of each of the viral genes and the Long Control Region (LCR) is drawn in relation to the virus genome.

1.5.1 Early Genes

There are six early genes that function to maintain the viral genome and promote cellular proliferation and viral DNA replication. Traditionally, the first genes to be expressed after virus infection are thought to be E6 and E7 which together function to promote cell growth and inhibit apoptosis. These genes will be discussed in more detail later. However, in 2002 a report was published showing that for HPV31b at least, E1 and E2 are the first viral genes to be expressed (Ozbun, 2002). E1 and E2 transcripts could be detected four hours after initial HPV31b infection of the keratinocytes and the other viral transcripts could not be detected until eight hours after initial infection (Ozbun, 2002). It was also suggested that the E1 and E2 transcripts originated from a different promoter than the major early promoter (Ozbun, 2002). Both of these scenarios are possible as the expression of E1 and E2 first would then control the expression of the other viral genes. However expression of E6 and E7 first is possible if expression of these genes can be activated by host transcription factors.

E1 is the viral helicase and binds to the origin of replication to promote viral DNA replication (Hughes & Romanos, 1993). E1 unwinds the viral DNA allowing the entry of the DNA replication machinery. E1 also actively recruits Topoisomerase I, DNA polymerase α and replication protein A (RPA) ready for DNA replication (Park et al., 1994, Loo & Melendy, 2004, Clower et al., 2006). E1 binding to the long control region is a weak interaction therefore E2 acts to aid the binding (Frattini & Laimins, 1994).

E2 is the viral replication and transcription factor and binds to the viral long control region regulating viral RNA transcription. It is conventionally thought that E2 controls the levels of E6 and E7 during the viral life cycle by binding to the four E2 binding sites on the HPV16 LCR and repressing E6/E7 expression and repression is lost when the viral genome integrates leading to the overexpression of E6 and E7 (Choo et al., 1987). However there may be other mechanisms controlling the overexpression of E6 and E7 for example involvement of host transcription factors. E2 has also been shown to bind to mitotic chromosomes through an interaction with a cellular protein and therefore allows the equal division of viral episomes between daughter cells during mitosis. For HPV16 the cellular partner is thought to be TopBP1 (Donaldson et al., 2007) and Brd4 is thought to be the cellular partner in BPV infections (You et al., 2004). BPV has also been shown to bind ChIR1 and tether to mitotic chromosomes (Parish et al., 2006a). Not all alpha-papillomaviruses interact with Brd4 and are still found tethered to mitotic chromosomes, therefore perhaps different viruses use different tethering proteins (McPhillips et al., 2006). E2 has also shown to bind to numerous nuclear proteins including p53, p300/CBP and SR proteins (Lee et al., 2000a, Parish et al., 2006b, Mole et al., 2009b). E2 has been shown to transcriptionally upregulate certain splicing regulatory SR proteins by binding to the promoters and activating transcription (Mole et al., 2009a, Mole et al., 2009b). E2 regulation has been shown for SRSF1, SRSF2 and SRSF3 (Mole et al., 2009a, Mole et al., 2009b). The upregulation of these splicing factors is an extremely important process as the viral life cycle depends on extensive splicing of viral RNAs therefore these splicing regulators are required for the processing of the viral RNAs, this will be discussed in more detail later.

E5 is expressed late in infection and is considered to be the major viral oncoprotein in Bovine papillomavirus infections. BPV E5 is required for transformation of BPV infected fibroblasts; however this is not the case for high risk HPV E5 (Yang et al., 1985) and in fact ß-HPVs do not encode E5 at all. The main function attributed to HPV E5 is the down regulation of MHC class I and II thereby inhibiting any immune response to viral infection (Zhang et al., 2003, Ashrafi et al., 2005). E5 has been shown to bind to MHC Class I and sequester the protein in the Golgi apparatus therefore preventing the MHC molecule from presenting viral peptides to the immune system for action (Ashrafi et al., 2005). E5 also interferes with the turnover of the Epidermal growth factor receptor (EGFR) leading to an increased number of receptors trafficked to the cell surface (Straight et al., 1993). The cells then have an exaggerated response to EGF and cellular proliferation is increased (Pim et al., 1992). E5 can bind to the 16kDa subunit of the vacular ATPase, which in turn alters the pH of the endosome (Rodriguez et al., 2000). The change in pH leads to altered trafficking of the endosomes. Therefore the expression of E5 late in infection will increase the proliferation of the normally replication inactive differentiating cells.

The final gene to be expressed is E4. Although E4 is technically termed an early gene it is in fact expressed later in infection as the cells differentiate. E4 does not have a start codon so the protein is expressed from a spliced isoform RNA encoding the first five amino acids of the E1 open reading frame to give a hybrid E1^E4 protein. E1^E4 binds to cytokeratins in the cytoplasm of the cell altering the cell structure and collapsing the whole keratin network, making viral escape easier by disrupting the integrity of the membrane (Wang et al., 2004). E1^E4 can also induce a G_2 arrest in infected cells allowing the viral DNA to continue to be replicated without the cell undergoing mitosis (Davy et al., 2002). E1^E4

binds cyclin B1 /cdk1 complex in the cytoplasm and prevents the complex from progressing the cell into mitosis (Davy et al., 2005). However this control must be overcome as the clinical definition of a wart is the presence of mitotic cells outwith the normal basal layer.

1.5.2 Late Genes

PVs express two late genes, L1 and L2. These proteins are the viral structural proteins and together they form the viral capsid. L1 is the major capsid protein and L2 the minor. They are only expressed in the granular layer of the epithelium to avoid triggering an immune response. Papillomaviruses have capsids composed of 72 capsomeres containing 360 copies of the L1 protein and around 12 copies of the L2 protein resulting in a 30:1 ratio of L1:L2 proteins per capsid (Greenstone et al., 1998). L1 is thought to be the major receptor binding protein. In 2009 an interesting report was published which may go some way to making sense of the cellular binding receptor controversy. Kines et al. (2009) showed that L1 bound to heparan sulphate proteoglycans on the basement membrane, which had been exposed after wounding the epithelium (Kines et al., 2009). Here the virus remained bound for an extended period of time. It was suggested that binding to the basement membrane initially allowed the virus to specifically infect proliferating cells, as cells that migrate to fill the wound site will divide to fill the layers of the epithelium (Kines et al., 2009). If the virus bound initially to epithelial cells out with the basement membrane, they would not be dividing and therefore the viral replication strategy would fail. This is in contrast to other reports that suggest PVs persistently infect epithelial stem

cells. A possible explanation for this discrepancy could be that the virus infects dividing cells and then changes the phenotype of the cell to be stem cell like in order for viral persistence. While bound to the basement membrane, the virus undergoes a conformational change resulting in L2 being cleaved and a neutralising epitope is exposed (Kines et al., 2009). It is thought that this conformational change exposes the cellular binding epitope of L1 and allows the virus to attach to the wound-invading basal cells (Kines et al., 2009). The conformational cleavage is thought to be carried out by furin which is secreted from the wounded epithelium or PC5/6 which is associated with the heparan sulphate proteoglycans (Kines et al., 2009). The authors suggested that it is the difference between the extracellular matrix of cultured cells and the basement membrane of the epithelium that may explain some of the different reports in receptor binding experiments (Kines et al., 2009).

1.5.3 E6

E6 and E7 cooperate in lesions and cancers to promote cell growth and avoid cellular apoptosis. For this reason these two proteins may be considered to be the viral oncoproteins for high risk papillomaviruses. E6 has three main methods of binding to protein partners; E6 contains two zinc finger motifs and a PDZ binding domain. The zinc finger domains are responsible for binding to E6's associated cellular binding partner E6AP. This interaction of E6 with E6AP can then go on to lead to numerous other protein interactions. E6AP is a cellular ubiquitin ligase that binds to E6 and therefore controls its function (Huibregtse et al., 1991, Scheffner et al., 1993). E6 has multiple cellular targets which it

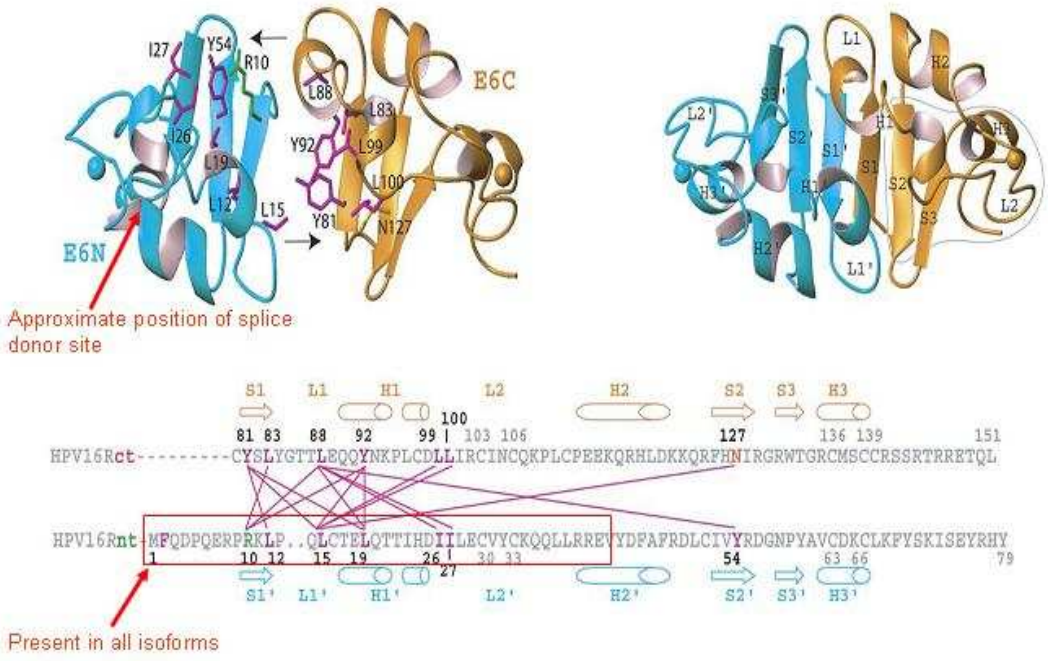


Figure 1.5 Crystal structure of the HPV16 E6 protein. The E6 protein contains two zinc finger domains and a PDZ-binding domain located at the extreme C-terminus of the protein. The N-terminal portion of the protein is shown in blue and the C-terminal portion in yellow adapted from (Nomine et al., 2006). The amino acids present in all isoforms are highlighted in the red box and the approximate position of the splice junction indicated by a red arrow on the crystal structure.

interacts with in order to alter the cell to allow for efficient viral infection and replication. The crystal structure of HPV16 E6 is shown in Figure 1.5. The E6 interacting partners are outlined in Table 1.

1.5.3.1 Avoidance of apoptosis

Perhaps the most infamous target for high risk E6 is p53 (Crook et al., 1991). p53 is a cellular tumour suppressor protein which acts to monitor the cell cycle and arrest the cell if there is any damage to the DNA before signalling for the cell to enter apoptosis by increasing the expression of the Bcl-2 family of apoptosis regulators (Schuler & Green, 2005). One of the functions of high risk E7 proteins is to degrade the retinoblastoma protein (p105) and other pocket proteins (p130 and p107) thereby driving the cell into S phase by relieving repression of the E2F transcription factor (Dyson et al., 1989). Under normal circumstances uncontrolled proliferation would result in the activation of p53 and apoptosis would occur. However E6 acts to circumvent this signal for apoptosis by degrading p53. This allows the infected cell to continue to proliferate and divide where it should not, thereby allowing viral replication to continue. E6 interacts with E6AP and alters its affinity for p53 (Hengstermann et al., 2001). p53 becomes extensively ubiquitinated and therefore degraded by the proteasome (Scheffner et al., 1993). Amino acid residue 47 has been shown to be partly responsible for directing the degradation of p53 as mutations in this amino acid lose the ability to degrade p53 (Nomine et al., 2006). However, all of the mutations studied by Nomine et al., (2006) retained the ability to bind E6AP suggesting there are other factors that are recruited to ubiquitinate p53

Binding Domain	Protein	Function
Via E6AP	p53	Apoptosis avoidance
	Bak	Apoptosis avoidance
	c-Myc	Apoptosis avoidance
	Tuberin	Proliferation
	hMCM7	Genome Instability
PDZ	hDlg	Loss of cell-cell communication
	hScrib	Loss of cell-cell communication
	MAGI 1-3	Loss of cell-cell communication
Direct binding via LXXLL motif	IRF-3	Immune System Activation
	E6AP	Ubiquitination
	Paxillin	Epithelial organisation
Unknown binding domain	Procaspase-8	Apoptosis Avoidance

Table 1: Table listing the E6 interaction partners with their binding domains and the consequences of the interaction listed.

and that mutation of amino acid 47 is interfereing with the recruitment of additional ubiquitin ligases.

High risk E6 also targets Bak for degradation. E6/E6AP can bind to and signal Bak for proteasomal degradation (Thomas & Banks, 1998). Bak is a member of the Bcl-2 group of apoptotic proteins, which is very similar to Bax, another proapoptotic member of the Bcl-2 family (reviewed in (Ola et al., 2011) . Bak and Bax have been shown to have very similar proapoptotic functions and there may in fact be some redundancy between their apoptotic functions. Bak, unlike Bax, has been found to be most active in the upper epidermal layers of the skin, the same location that HPV replication occurs (Krajewski et al., 1996, Thomas & Banks, 1998). It is thought that under normal conditions within the cell, E6AP regulates the levels of Bak, however the presence of HPV18 E6 has been shown to decrease Bak levels through proteasomal degradation (Thomas & Banks, 1998). This degradation of Bak prevents the protein being able to permeabilise the mitochondrial and ER membranes, therefore also avoiding the activation of the caspase cascade and preventing the cell from entering apoptosis (Adams & Cory, 2007). E6/E6AP has not been shown to target Bax in the same manner as it targets Bak, however this may be because p53 has been shown to regulate Bax within the cell (Moll et al., 2005). The degradation of Bak interferes with the cells ability to enter apoptosis through activation of the intrinsic apoptosis pathway (Adams & Cory, 2007). c-Myc is another pro-apoptotic protein that is targeted for degradation by E6/E6AP (Gross-Mesilaty et al., 1998). c-Myc is a cellular protein that has two seemingly conflicting functions. c-Myc controls both cellular proliferation and is also involved in the apoptotic pathway (reviewed in (Soucek & Evan, 2010). Under normal circumstances low level expression of c-Myc within quiescent cells drives the cells into S-phase and

inhibition induces a growth arrest, therefore controlling cellular proliferation (Pucci et al., 2000). However overexpression of c-Myc drives the cell into apoptosis (Pucci et al., 2000). E6/E6AP binds to c-Myc and results in its degradation (Gross-Mesilaty et al., 1998). The degradation of c-Myc may seem strange due to the proliferation properties of c-Myc; however this degradation is probably to alleviate the proapoptotic functions of c-Myc, as E6 also results in the degradation of other apoptotic proteins.

1.5.3.2 PDZ domain-containing proteins

In order to be able to transform the infected cells, HPV must be able to disrupt normal inter- and intra- cellular communication pathways. E6 contains a PDZ binding domain at the extreme C-terminus of the protein; in fact the last four amino acids of high risk E6 proteins comprise a PDZ binding domain: X-(S/T)-X-(V/I/L). HPV16 E6 ends with E-T-Q-L and HPV18 ends with E-T-Q-V. Proteins with this motif at their C terminus can bind to partner proteins containing PDZ structural domains which are around 90 amino acids in length (Cho et al., 1992). Low risk papillomavirus E6 proteins do not end with a PDZ binding domain which has led to the hypothesis that the E6-PDZ domain interactions are important for transformation of cells infected with high risk HPV (Kiyono et al., 1997). E6 interacts with all of the following proteins, which are involved in the maintenance of tight junctions between cells, via an interaction with their PDZ domains. PDZ domains are found mainly in signalling molecules with roles in cell polarity and differentiation. These so-called scaffold proteins have been proposed to mediate assembly of protein complexes at key cellular sites and to integrate cell signalling. hScrib, hDlg and MAGI 1-3 are all members of this

group of proteins. Interestingly, E6 proteins from different high risk viruses have differing specificities for degradation of different PDZ proteins. For example HPV16 E6 directs the degradation of hScrib preferentially over hDlg whereas HPV18 E6 does the opposite (Thomas et al., 2005).

hScrib is the human homologue of the *Drosophila* Scribble protein which normally regulates the growth of the cells and can also act as a tumour suppressor (Dow et al., 2003). hScrib is normally located at tight junctions between cells. E6/E6AP binds to hScrib and facilitates its degradation therefore reducing the control over cellular growth and also results in the cell loosing contact with the extracellular matrix and therefore allowing cellular transformation and tumour invasion to occur (Nakagawa & Huibregtse, 2000).

hDlg is the human homologue of the Dlg tumour suppressor in *Drosophila*. Like hScrib it is involved in epithelial tight junctions and is also involved in cellular adhesion, cellular proliferation and cellular polarity (Woods & Bryant, 1991, Woods et al., 1996). hDlg also acts as a scaffold protein allowing a platform for protein interactions (Humbert et al., 2003). Again E6 binds to hDlg and results in the ubiquitin-mediated degradation of hDlg (Gardiol et al., 1999). This results in a loss of cellular adhesion, proliferation and polarity leading to tumourigenesis. The staining pattern of hDlg in transformed cells is very different to that of non-transformed cells. In normal epithelial cells hDlg is found predominately at tight junction plaques between neighbouring cells, however in transformed cells hDlg staining is predominately cytoplasmic suggesting a loss of cell-cell communication between neighbouring transformed cells. As E6 has been shown to bind to PDZ domain containing proteins it is thought that E6 is the protein which is involved in the redistribution of hDlg during transformation.

MAGI-1-3 are also involved in the regulation of tight junctions (Laura et al., 2002). Interaction with these proteins will disrupt adhesion to the extracellular matrix. MAGI proteins are degraded by ubiquitin-mediated proteasomal degradation directed by E6 (Glaunsinger et al., 2000, Thomas et al., 2002). MUPP1 is another E6 protein partner, which controls cellular proliferation and is also involved in control of cell junctions and cell polarity (Lee et al., 2000b). Therefore when E6 binds to MUPP1 cell growth becomes uncontrolled and the cells loose structural integrity.

All of the previously mentioned functions for E6 degradation of PDZ domain containing proteins have been highlighted to explain transformation of the HPV infected cells. While these functions are active when E6 becomes overexpressed (e.g. when the HPV genome becomes integrated and the cells are transformed), a recent study has also shown that PDZ domain-containing proteins can bind to and stabilise low levels of E6 expressed during an infection aiding in maintaining the viral episomes. Nicolaides et al (2011) showed that the PDZ binding domain of E6 is required for E6 stability (Nicolaides et al., 2011). hScrib, and other PDZ domain containing proteins, can bind to and stabilise HPV16 E6 levels in transfected epithelial cells and this stabilisation is dependent on the PDZ domain of the hScrib and the PDZ binding domain of E6 (Nicolaides et al., 2011). The stabilisation of E6 led to enhanced episome maintenance as a mutated interaction between hScrib and E6 resulted in frequent loss or integration of the HPV genomes (Nicolaides et al., 2011). Therefore this PDZ domain interaction for E6 is important for viral infection as well as for transformation.

1.5.3.3 Transcriptional Regulators

E6 can not only inhibit apoptosis and disrupt communications between neighbouring cells; E6 can also disrupt transcription of various genes through binding to transcriptional regulators. Perhaps the most well known transcriptional regulator targeted by E6 is p300/CBP (Zimmermann et al., 1999). p300/CBP is a transcriptional co-activator of p53, therefore not only does E6 act to degrade p53 directly, it also inhibits p53 action by preventing p300/CBP activation. The inability of p300/CBP to bind and activate p53 will reduce p53's ability to activate the DNA damage response (Kessis et al., 1993). E6 binds to p300/CBP via its second zinc finger domain (Zimmermann et al., 1999). Similarly, E6 also promotes the degradation of Tuberin which relieves the control over pro-proliferation genes normally exerted by Tuberin (Lu et al., 2004). Tuberin along with its binding partner Hamartin normally inhibits S6 kinase and thereby prevents the activation of eIF4E-binding protein 1 (4EBP1) (Lu et al., 2004). The inactivation of 4EBP1 suppresses protein translation which negatively controls cellular proliferation. E6 induced degradation of Tuberin prevents it from binding Hamartin and therefore indirectly activates protein translation and cellular proliferation.

1.5.3.4 Immune Recognition Regulation

All viruses must by nature avoid the immune response. HPV can establish a persistent infection therefore it must have evolved ways of avoiding recognition by the immune system. HPV16 E6 has been shown to bind to Interferon Regulatory Factor-3 (IRF-3) and prevents its transcriptional activity (Ronco et

al., 1998) . IRF-3 has two modes of antiviral activity, it can either directly activate Bax leading to apoptosis or it can transcriptionally activate antiviral genes as part of a virus activated transcription factor complex (Wathelet et al., 1998). Interference with this antiviral pathway is common to many viruses including RNA and DNA viruses (Chattopadhyay et al., 2011). Under normal circumstances after detection of a viral infection the cell will undergo apoptosis, IRF-3 normally acts to control apoptosis progression. IRF-3 as part of the larger transcription factor complex can transactivate the IFN-β promoter, which therefore means that by disturbing IRF-3 functions, HPV16 can avoid the activation of the antiviral response. Ronco et al (1998) showed that IRF-3 can transactivate the promoter of a reporter construct and co-transfecting HPV16 E6 prevented this transactivation while cotransfection of HPV6 E6 did not. This suggests that high risk viruses have evolved to inhibit the antiviral immune response.

Not only does HPV16 E6 prevent IRF-3 activity but it also prevents expression of TLR9. Toll Like Receptors are cellular receptors that recognise specific molecular patterns and alters the immune system to an infected cell. TLR9 is activated in response to foreign double stranded DNA. HPV16 E6 and E7 can down regulate expression of TLR9 mRNA therefore allowing the virus to escape immune recognition (Hasan et al., 2007). HPV16 oncoproteins were shown to be much more efficient at reducing TLR9 mRNAs than those from HPV18 or HPV6 (Hasan et al., 2007) which could be one of the reasons why HPV16 is much more efficient at establishing a persistent infection and is more frequently associated with cervical cancers.

1.5.3.5 Epithelial Organisation and Differentiation

HPV16 E6 has been shown to bind to paxillin and disrupt the actin fibres that form the cytoskeleton (Tong & Howley, 1997). Paxillin plays an important role in the structural organisation of the cell. When epithelial cells were transfected with BPV1 E6 the actin filaments were severely disrupted (Tong & Howley, 1997). HPV16 E6 has also been shown to bind to ERC-55 or E6-BP which is a calcium binding protein (Chen et al., 1995). The interaction with E6-BP is likely to disrupt the normal calcium signalling in the cell thereby postponing epithelial differentiation and promoting mitosis (Chen et al., 1995).

1.5.3.6 Genome Stability

An important feature of transformed cells is as they become more transformed their genomes become more unstable, allowing mutations to build up and contributing to the transformed phenotype of the cells. HPV16 E6 has been shown to bind to the pRb binding domain of hMCM7 (human minichromosome maintenance seven protein) (Kukimoto et al., 1998). Normally hMCM7 and other members of the hMCM family bind to chromatin before S phase entry and are displaced during DNA replication (Fujita et al., 1996). If hMCM7 does not reattach to the chromatin then only one round of DNA replication can take place. pRb normally binds to the C-terminus of the hMCM7 protein, which inhibits DNA replication in Xenopus (Sterner et al., 1998). Therefore Kukimoto et al (1998) suggested E6 binding to hMCM7 prevents pRb binding and thus relieves repression of DNA replication. This deregulated DNA replication may allow genetic mutations to arise and build up, leading to genomic instability.

Not only does HPV16 E6 allow mutations to build up it also acts to prevent repair of single strand DNA breaks. HPV16 E6, HPV8 E6 and HPV1 E6 can all bind to and inhibit XRCC1 thereby preventing repair of single strand breaks in the host genome allowing mutations to arise and not be repaired (Iftner et al., 2002). These authors suggested a possible benefit to the virus in that XRCC1 normally interacts with DNA polymerase ß and by interacting with XRCC1, E6 is releasing DNA polymerase ß and promoting DNA replication (Iftner et al., 2002). This hypothesis makes sense as above all viruses want to replicate and produce progeny viruses. There is no benefit to the virus in causing cancers. Similarly HPV16 E6 has been shown to induce the proteasomal degradation of O(6)-methylguanine DNA methyltransferase preventing the DNA repair functions normally exerted by this molecule (Srivenugopal & Ali-Osman, 2002).

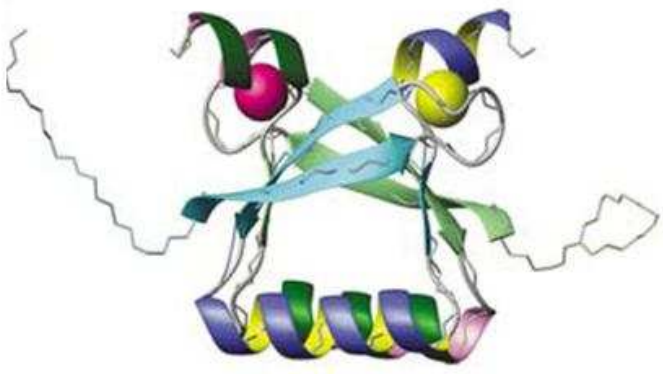


Figure 1.6 Crystal structure of the HPV45 E7 protein. The E7 protein contains a zinc finger domain and three CR domains. The Conserved Region domains are named due to their homology to Adenovirus E1A pRb binding domains. The E7 proein is shown as a dimmer with zinc atoms indicated in red and yellow. Different dimmers are shown in green and blue (Ohlenschlager et al., 2006). The highly structured C-terminus of E7 is present in the E6*X isoform.

1.5.4 E7

The E7 protein is made up of three conserved regions; Conserved Region (CR) 1, 2 and 3. The crystal structure of HPV16 E7 is shown in Figure 1.6. HPV45 E7 is composed of 3 beta sheets and 2 alpha helices. E7 has homology with SV40 Large T antigen and adenovirus E1A protein (Dyson et al., 1989). The conserved regions are named due to their high homology with adenovirus E1A which also binds to and degrades pRb. The most well known interaction of E7 is with the retinoblastoma family of proteins including the pocket proteins pRb (p105), p107 and p130 (Dyson et al., 1989). The N-terminus of E7 is unstructed and flexible meaning the CR2 domain responsible for pRb binding is highly accessible (Ohlenschlager et al., 2006). E7 binds to pRb inducing its degradation (Boyer et al., 1996). This degradation releases pRb's association with E2F (Chellappan et al., 1992). E2F is a transcriptional activator which is involved in regulation of transcription of genes encoding proteins essential for the progression of the cell cycle (reviewed by (Sun et al., 2007). The release of E2F from pRb allows E2F to bind to and activate the promoter regions of many S phase and cellular growth-related genes including cyclin A and cyclin E which are both required to allow the cell to enter S phase (Melillo et al., 1994, Sun et al., 2007). Under normal circumstances if the cell has encountered damage or circumstances that are not permissive for replication, pRb binds to E2F and prevents it binding to and activating promoters of S phase genes. However, E7 degradation of pRb results in the release of E2F and therefore the expression of S phase entry genes resulting in the replication of cellular and viral DNA.

Not only is cyclin A and cyclin E expression increased due to E2F activation, HPV16 E7 has also been shown to directly bind to these cyclins independent of

pRb binding. HPV16 E7 has also been reported to bind and inhibit p21 and p27 cyclin inhibitory functions (Zerfass-Thome et al., 1996, Funk et al., 1997, Jones et al., 1997). p16 and p21 are associated with cell cycle exit in keratinocytes leading to differentiation. Therefore the inhibition of these proteins is important for viral DNA replication in cells which would normally be replication inactive. Furthermore E7 binds E2F6 and prevents its normally repressive functions on E2F activated S-phase genes (McLaughlin-Drubin et al., 2008). E2F6 is a member of the polycomb repressor complex that normally associates with chromatin and prevents transcription (McLaughlin-Drubin et al., 2008). It was suggested by the authors that this repression of E2F6 activity resulted in an extended S-phase allowing viral DNA synthesis to continue (McLaughlin-Drubin et al., 2008).

E7 also interacts with other transcription factors including p300/CBP, PCAF and Mi2ß. p300/CBP are transcriptional co-activators and are involved in many pathways in the cell, including DNA repair, cell proliferation and apoptosis (for review see Goodman and Smolik., 2000). p300/CBP can activate transcription of proteins which control the cell cycle making them important tumour suppressor proteins (for review see Goodman & Smolik., 2000). Bernat et al. (2003) showed that E7 interacts directly with p300 and results in a decrease in its transcriptional activity (Bernat et al., 2003).

E7 has also been shown to interact with p600 a retinoblastoma associated protein through the N-terminal domain of E7 (Huh et al., 2005). This interaction has been shown to be independent of pRb (Huh et al., 2005). The interaction of E7 with p600 is thought to be important for the transforming activities of HPV16 E7 (Huh et al., 2005). Reduction of p600 in transformed cells results in a

reduction in anchorage-independent growth whether the cells are HPV-positive or not (Huh et al., 2005). It has also been shown that p600 is an important regulator of anoikis. Anoikis is a form of apoptosis which is initiated when cells are stimulated to grow without contact with a matrix (for review see Chiarugi & Giannoni., 2008). For example in differentiating epithelia, cells should not undergo cell division when they are no longer attached to the basement membrane. Anoikis is the form of apoptosis which prevents this form of cell division from occurring. In papillomavirus infections, where there is division outside of the basal layer, this control pathway must be overcome. Therefore E7 binding to p600 is an important step to allow viral DNA synthesis during infection and also once the cells have become transformed and are growing in an uncontrolled manner.

E7, along with E6, contributes to chromosomal instability by binding to γ-tubulin and altering the recruitment of γ-tubulin to centrosomes (Nguyen et al., 2007). Interfering with centrosomes indirectly affects the ability of the chromosomes to segregate properly and divide evenly between daughter cells during mitosis. During a normal mitosis, the cell has exactly two centrosomes; however one of the hallmarks of HPV infection is the presence of multipolar centrosomes (Duensing & Munger, 2003, Duensing et al., 2008). The main action of E7 is through uncoupling the synthesis of new centrosomes from the cell cycle by degrading pRb/p107/p130 and interfering with cdk2 activity (Dyson et al., 1989, Nguyen & Munger, 2008). However Nguyuen et al. (2007) showed that E7 can also bind to γ-tubulin and retard its association with centrosomes, thus interfering with the formation of new centrosomes (Nguyen et al., 2007). The build up of genetic instability in HPV-infected cells is also likely added to by the interaction of E7 with Nuclear Mitotic Apparatus Protein 1 (NuMA) and dynein

(Nguyen & Munger, 2009). NuMA and dynein are both found at spindle points during mitosis and when E7 is expressed in the cell, these proteins are delocalised resulting in errors during mitosis (Nguyen & Munger, 2009). NuMA is targeted to the spindle poles by dynein and acts to stabilise the fibres (Nguyen & Munger, 2009). NuMA and dynein importantly have also been shown to be involved with cell polarity organisation (Nguyen & Munger, 2009). NuMA has been shown to localise at the apical membrane of dividing basal cells, identifying the cell which is to go on to further divide from the cell which has committed to differentiation. Dynein has also been reported to have similar functions (Nguyen & Munger, 2009). Nguyen and Münger. (2009) reported that E7 interacts with NuMA and interferes with the protein's ability to bind to protein partners, including dynein, thereby inhibiting the formation of the spindle poles. However, as this is a dynamic process some active NuMA and dynein would eventually reach the spindle pole, thereby delaying mitosis and allowing the build up of mitotic abnormalities (Nguyen & Munger, 2009).

Genomic instability is a consequence of transformation by high risk viruses. The uncontrolled proliferation induced by HPV16 E7 and the lack of an apoptotic pathway as a result of E6 activity allows the build up of genetic mutations and E7's interference with mitosis only adds to this. However this is unlikely to be the sole reason for E7 interaction with NuMA as low risk E7 proteins also bind NuMA (Nguyen & Munger, 2009). Nguyen and Münger. (2009) also suggested that this interaction could be required for viral life cycle in that disruption of NuMA will also interfere with the identification of cell polarity, therefore it may allow viral DNA replication in cells which have moved into the suprabasal layer and should be entering differentiation (Nguyen & Munger, 2009).

1.6 Transformation

E6 and E7 are two of the earliest genes expressed after viral infection and the RNAs encoding them are expressed throughout the epidermal layers of the skin. E6 and E7 are the two major oncogenes of the high risk HPVs, which along with E5 account for the oncogenicity of these viruses. There is convincing evidence of the role of E5 in the transformation process during a BPV viral infection, however the role of alpha virus E5 during the transformation process is not quite so convincing. In fact, after genome integration, E5 is no longer expressed. So E5 is probably not present in many cervical cancers. E6 and E7 both have many cellular interacting partners that are regulated in order to allow increased cellular proliferation and cause cell transformation. In a normal infection, E6 and E7 are expressed at low levels acting to allow viral replication in response to epithelial differentiation. During cell transformation, the genome breaks into a linear form and becomes integrated into the host cell chromosome and the repression of the levels of E6/E7 expression is disrupted resulting in E6 and E7 being overexpressed (Choo et al., 1987). This disruption is due to the break in the circular genome usually occurring around the E1 and E2 region of the genome (Choo et al., 1987).

E2 in combination with its other functions, discussed above, represses the expression of E6 and E7 in a normal infection (Bernard et al., 1989, Tan et al., 1992, Tan et al., 1994). Disruption in the E2 gene leads to deregulated expression of E6 and E7. While E6 and E7 expression at low levels is necessary for the productive viral life cycle, high levels of expression lead to chromosome abnormalities and genomic instability which then aids in the transformation of the infected cells. The proliferative effects of E7 lead to uncontrolled DNA

replication allowing mutations to build up in the cellular DNA. Likewise E7's ability to interfere with mitotic spindle and centrosome assembly adds to the genomic instability (Nguyen & Munger, 2009). E6's ability to block apoptotic pathways allows the cells to continue with their cell cycle even though there have been mistakes made thereby building upon E7's effects (Crook et al., 1991, Thomas & Banks, 1998). E6 and E7 expression is sufficient to immortalise cells but other changes in the cell are required to malignantly transform the cell.

1.7 Cell Culture Model System

Due to the complexities of the HPV viral life cycle and its tight link with epithelial differentiation the virus is difficult to study in the laboratory environment. Until fairly recently the only methods of studying HPV gene action were by either transiently or stably expressing virus proteins in transformed cells. While this is good enough for binding studies, protein/protein interactions and for functional studies on the individually expressed proteins it does not allow any functional information from the natural system where multiple viral proteins are expressed at any one time in a background of changing cellular protein expression. In order to get an idea of the functions of the viral proteins at various stages in the virus life cycle, a system that allows for differentiation of infected epithelial cells is required.

There are three cell culture systems that are used when studying the papillomavirus life cycle; namely organotypic raft culture systems; HPV stably transformed NIKS cells and W12 cells. Raft cultures are systems that provide cells with a 3D support system allowing the cells to fully differentiate in culture.

The epidermal cells are grown on a matrix of collagen on top of a wire grid at the liquid air interface, where any nutrients must diffuse to the upper layers via a concentration gradient similar to the *in vivo* system. The rafts are grown for around two weeks to allow the cells to differentiate. If the epithelial cells contain HPV genomes they can produce infectious virus. Raft cultures are excellent systems for studying the complete virus life cycle or late viral proteins that require differentiation for production. However, they are relatively time consuming and cumbersome to handle in the laboratory. While these systems are the best for inducing complete differentiation and efficient virus production, other less cumbersome systems are often routinely used where almost complete differentiation of the cells and less efficient virus production is acceptable.

One such system for studying the virus life cycle is the NIKS system. NIKS stands for Normal Immortal KeratinocyteS and were originally isolated by Professor Allen-Hoffmann in 1999 from the BC-1-Ep strain of human neonatal foreskin keratinocytes (Allen-Hoffmann et al., 2000). The NIKS cells spontaneously arose after passaging normal foreskin keratinocytes until senescence (Allen-Hoffmann et al., 2000). The cells that survived this crisis were termed NIKS. NIKS cells retain the ability to differentiate in culture similar to the parental cells (Allen-Hoffmann et al., 2000). They also have wild type p53 levels and do not undergo anchorage independent growth (Allen-Hoffmann et al., 2000). The major identifiable difference between the NIKS and the parental cells is the appearance of an extra copy of the long arm of chromosome 8 in the NIKS cells (Allen-Hoffmann et al., 2000). Significantly the c-Myc gene is located on chromosome 8, however steady state levels of c-Myc RNA in NIKS are not increased compared to the parental cells (Allen-Hoffmann et al., 2000). Importantly, an extra copy of chromosome 8 is often associated with increased

growth in human cells as an extra chromosome 8 is often found in cancers including breast and prostate cancers (Allen-Hoffmann et al., 2000). These NIKS cells were then stably transfected with HPV16 genomes taken from the W12 cell line (see below) and a GFP expression plasmid conferring G418 resistance (Flores et al., 1999). These cells were shown to fully differentiate in raft culture and therefore could complete a productive viral infection (Flores et al., 1999). The NIKS cells could not only support the HPV16 viral life cycle but also other HPV types such as HPV31b (Flores et al., 1999). The ability of these cells to support a productive virus infection has made them ideal for studying papillomavirus life cycles. Importantly Flores et al were able to detect virus like particles in the nucleus of the transfected NIKS in raft culture suggesting the virus can indeed undergo a productive virus life cycle (Flores et al., 1999). The only caveat for using these cells in my view is that the genomes have been transfected and NIKS have not been infected "naturally", however these cells can be passaged so will be continuously available whereas stocks of other naturally occurring cell lines, such as the W12 line, will eventually run out.

In our laboratory we use the W12 model system. W12 cells are cervical epithelial cells that were originally isolated from a low grade cervical lesion by Professor Margaret Stanley in 1989. W12E (clone 20863) cells contain around 100 episomal copies of the HPV16 genome and can differentiate in monolayer and raft culture allowing completion of a productive viral infection (Stanley et al., 1989, Jeon et al., 1995). As these cells were isolated from a low grade lesion, the HPV16 genomes are naturally occurring virus genomes and the cells have been immortalised by the virus itself making the system arguably much more like the natural infectious situation. The W12 system was evolved to produce a second cell line which contains integrated HPV16 genomes, called W12G cells

(clone 20861) (Jeon et al., 1995). W12E and W12G cells can both differentiate in culture and are not transformed. They require murine J2 3T3 feeder layer cells and mitogens including EGF, Cholera toxin and hydrocortisone for growth. In our laboratory two more cells lines were sequentially derived from W12G cells. W12GPX were derived by growing W12G cells without feeder layer cells. Following crisis and cell death, a new cell line, W12GPX was established that was able to grow in the absence of feeder cells. W12GPXY cells were then derived from W12GPX cells by growing them in the absence of mitogens in the medium (Aasen et al., 2003). W12GPX and W12GPXY cells display a transformed phenotype and are invasive in raft cultures. Moreover, when injected into nude mice, W12GPXY cells formed large squamous cell carcinomas (Aasen et al., 2003). W12GPX and W12GPXY cells have lost the ability to differentiate as when grown in organotypic raft culture there was no differentiation (Aasen et al., 2003). Finally, while W12GPX cells grow in monolayer culture in large colonies similar to untransformed epithelial cells, W12GPXY cells grow in small colonies. The W12 model system comprising these four cell lines allows the study of both the productive viral life cycle and the HPV16-induced transformation of the cervical epithelial cells. Like other well known cervical cancer cell lines W12GPX and W12GPXY cells only express the HPV16 E6 and E7 proteins, however the benefit of using the W12 cell lines is that we have a series of cell lines derived from a common ancestor cell with varying phenotypes relating to increased transformation with W12G cells being non-transformed and the W12GPXY cells are fully transformed with the W12GPX cells in an intermediate state of transformation. A flow chart showing the derivation of the W12 cell lines is shown in Figure 1.7.

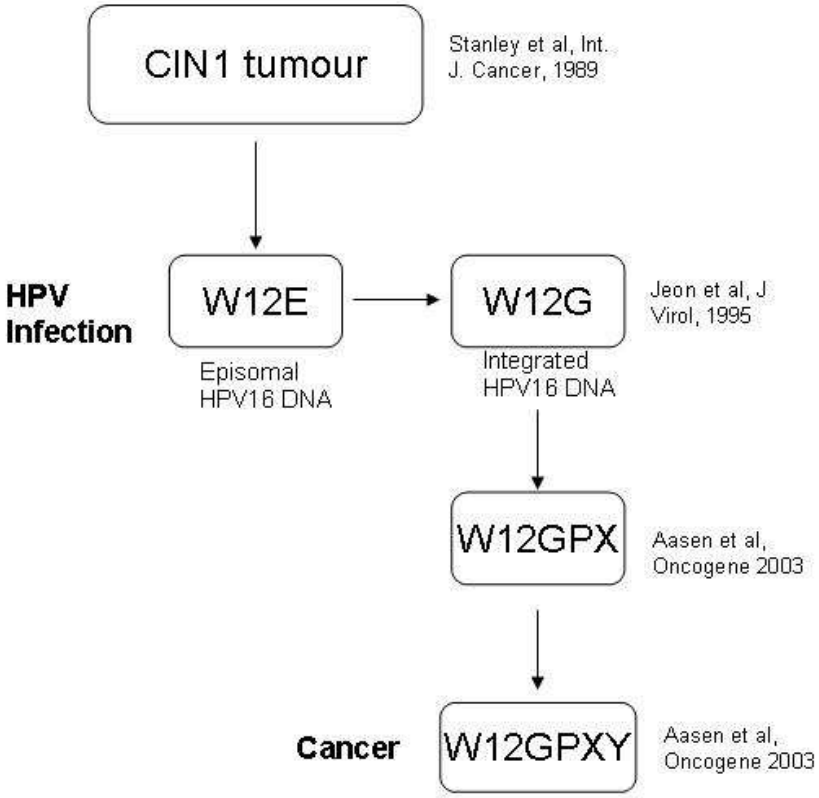


Figure 1.7 Flow chart of W12 cell line. W12E cells containing episomal copies of the HPV16 genome and W12G cells containing integrated copies of the HPV16 genome were isolated from a low grade CIN 1 lesion: both are immortalised but not transformed. W12E cells allow the study of the virus life cycle. W12GPX and W12GPXY cells were further derived from W12G cells and are transformed cervical epithelial cells. W12G, W12GPX and W12GPXY cells allow the study of HPV16 induced cellular transformation.

2 RNA Processing

RNA expression is a tightly regulated process that involves a number of steps and levels of control. There are a number of stages of processing before RNAs are considered to be mature and functionally active. RNAs must be transcribed, capped, spliced and polyadenylated before they can be exported from the nucleus and translated. Each stage is tightly regulated with a number of proteins involved at every step. It should also be said that these steps do not occur independently, capping, splicing and polyadenylation occur co-transcriptionally. RNA processing is one of the most important processes in the cell. Alterations in RNA processing can have wide ranging results as simply skipping an exon or including an intron in an mRNA can change the function of the resulting protein.

2.1 Capping

After transcription the first processing stage of the newly transcribed RNA is capping. The newly formed RNA will exit the RNA polymerase with a pppA 5' end (Moore & Proudfoot, 2009). Three enzymatic reactions then occur to convert the triphosphate 5' end structure (pppA) to 7 methyl Guanosine cap (7meGpppA) (reviewed by (Zorio & Bentley, 2004). Once recruited to the promoter region, the polymerase C-terminal domain (CTD) becomes active by phosphorylation of the serine 5, and this phosphorylation recruits the capping enzymes to the 5' end of the RNA (Hocine et al., 2010). Once capped the

phosphorylation status of the C-terminus of the polymerase changes and the serine 2 becomes phosphorylated which signals that transcription elongation is to take place (Hocine et al., 2010). As the newly formed RNA emerges from the RNA polymerase one of the triphosphates is removed by RNA terminal phosphatase allowing guanyl transferase to add a GTP to the chain (Hocine et al., 2010). Methyl transferase then methylates the guanine (Hocine et al., 2010). The cap structure provides stability to the freshly synthesised RNA preventing the degradation by exonucleases in the nucleus (Shimotohno et al., 1977). The cap also aids in the export of the RNA from the nucleus to the cytoplasm for translation by recognising the nuclear export complex (Hamm & Mattaj, 1990). Once added the cap is bound by the cap binding complex which remains bound until it is replaced by translation initiation factors. The cap is recognised by the translation initiation factor complex (eIF4F), comprising of eIF4E which recognises the cap structure, the RNA helicase eIF4A and eIF4G which forms a scaffold, thereby enhancing the translation of the RNA (Pierrat et al., 2007). Finally, it is also thought that the cap binding complex aids polyadenylation as the newly transcribed RNA forms a loop and the cap binding complex is found bound to poly(A) binding protein (PABP) that attaches to the polyadenylated tail of the mRNA (Cooke & Alwine, 1996).

2.2 Alternative and Constitutive Splicing

Splicing is the process of the removal of introns and joining of exons from premature mRNAs to form mature mRNAs. It is a very complex process with a number of steps and multiple proteins involved. There are two types of splicing:

constitutive and alternative splicing. Constitutive splicing is where particular introns are removed every time the RNA is expressed. Alternative splicing is where the removal of introns or exons can be different depending on conditions or signals within the cell. Alternative splicing increases the number of functionally different proteins that can be expressed from single RNAs. A diagram of constitutive and alternative splicing is shown in Figure 1.8. There are different types of alternative splicing, for example there is intron retention where the normally non-coding intron is retained within the RNA or alternative exon splicing is where in one RNA one exon is retained and another spliced out whereas in another RNA the retained exon from the first RNA is removed and the exon that was removed in the first RNA is retained. Around 94% of all human RNAs are alternatively spliced (Ward & Cooper, 2010).

The process of splicing is shown in Figure 1.9. The first step in the splicing process is recognition of the 5' splice site by U2 auxiliary factor U2AF (Graveley, 2000). U2AF binds to the pyrimidine tract and the 3' splice site and SF1/mBBP then locates and binds to the branch point sequence (reviewed by (Graveley, 2000). U1 small nuclear ribonucleoprotein (snRNP) then binds U2AF (reviewed by (Graveley, 2000). This is termed the E or early complex. U2 snRNP binds to the branch point sequence converting the E complex to the A complex (Graveley, 2000). B complex is then formed by U4/U6 and U5 snRNPs binding to the complex (reviewed by (Graveley, 2000). The complex is rearranged to form the C complex when U6 displaces U1 and interacts

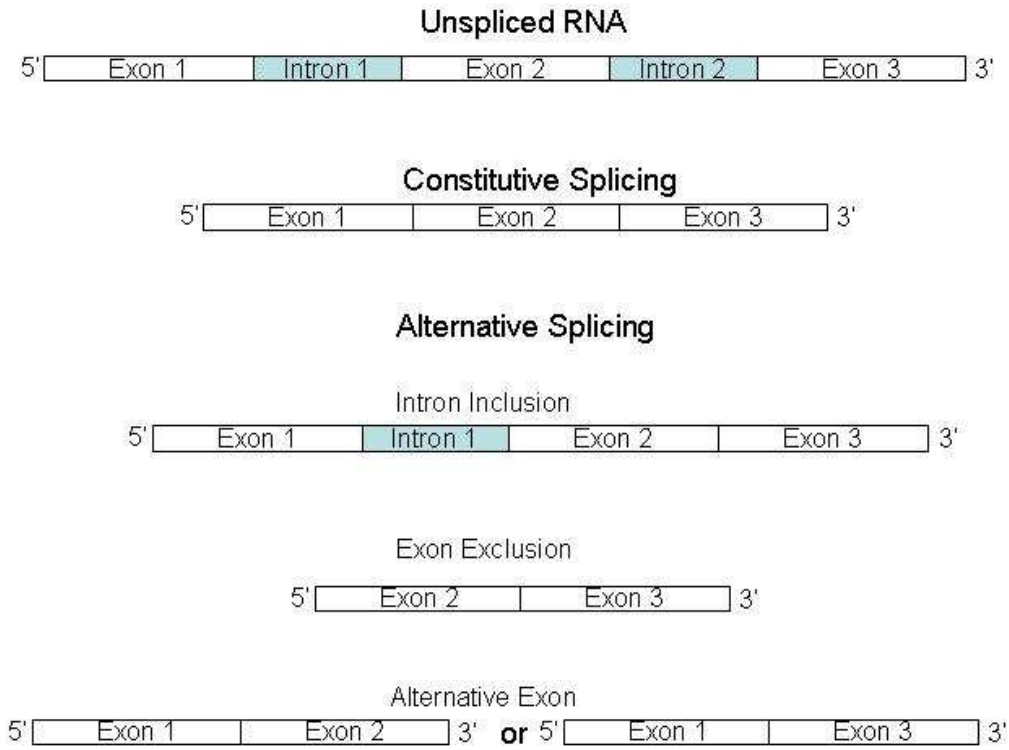


Figure 1.8 Diagram showing some examples of alternative splicing. Splicing can be divided into two types, Constitutive and Alternative splicing. Constitutive splicing is when introns are removed and exons are joined together every time the mRNA is expressed. Alternative splicing arises when under some conditions an intron can be retained in the resulting mRNA (Intron retention) or exons can be removed (exon exclusion) or exons can be included or removed depending on cellular signals (alternative exon).

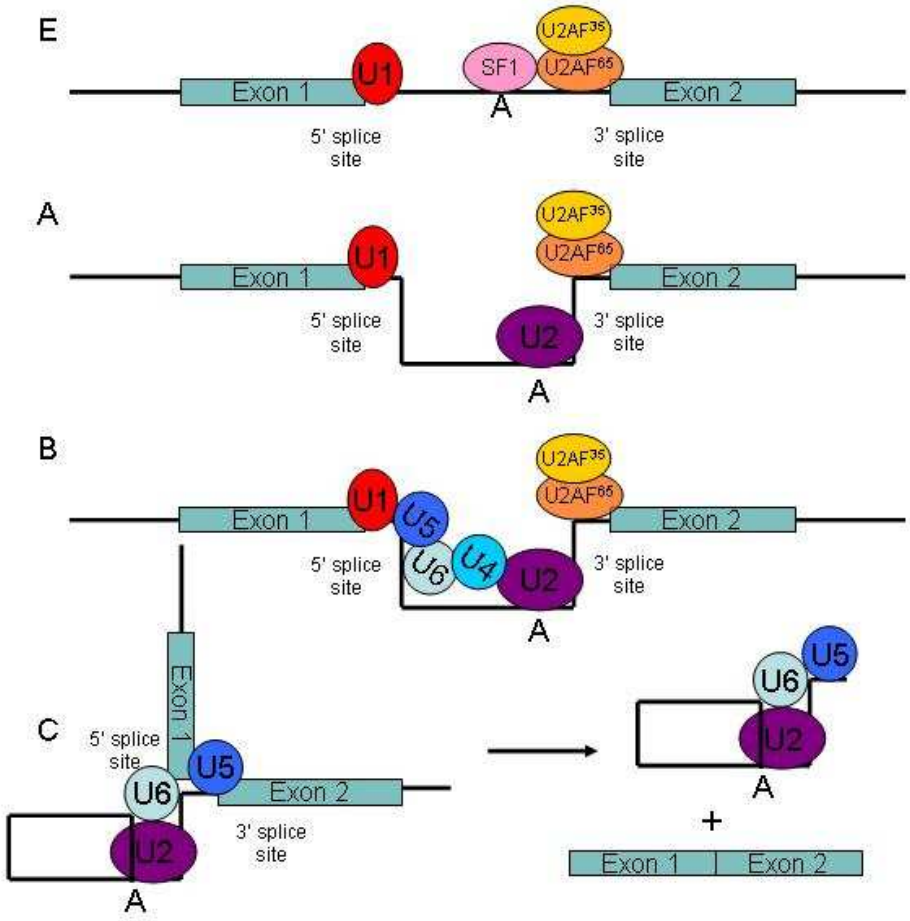


Figure 1.9 Schematic diagram of the splicing process. E) shows the E complex formation with the U1 snRNP (red) bound to the 5' splice site and U2AF (orange) bound to the 3' splice site and the SF1 (pink) bound to the branchpoint. A) The A complex then forms with U2 snRNP (purple) replacing the SF1 at the branchpoint. B) The B complex arises when the U6/U4 U5 complex binds to the RNA. C) The C complex where the actual splicing reaction occurs. The 5' splice site bound by U6 attacks the branchpoint forming a "lariat" structure. The 3' end of the first exon then cleaves the 5' end of the second exon and joins together. The lariat structure containing the intron is removed and degraded.

with U2 snRNP that is already bound to the RNA and form a loop which is now the active spliceosome and U5 snRNP stabilises the loop (Graveley, 2000). The C complex is the active splicing complex. The branch point A residue can form an unusual phosphodiester bond with the G residue of the 5' splice site forming a loop or "lariat" structure (Hocine et al., 2010). The 3' end of the cleaved exon 1 then cleaves the 5' splice site of the second exon and can join together to form a mature mRNA (Hocine et al., 2010). The intron is removed still in the lariat structure bound by snRNPs ready for degradation (Hocine et al., 2010). The whole process is controlled by serine/arginine rich (SR) and heterogeneous nuclear ribonucleoproteins (hnRNP) proteins. SR proteins act to enhance splicing by binding to sequences within the open reading frame called exonic splicing enhancers and intronic splicing enhancers and recruiting members of the spliceosome (Wu & Maniatis, 1993, Staknis & Reed, 1994, Shen et al., 2004, Cho et al., 2011). hnRNP A/B proteins bind to the exonic splicing suppressors or intronic splicing suppressors and act to inhibit splicing. SR and hnRNP proteins have antagonistic functions and the levels of the proteins and where on the pre-mRNA they bind determine which splice sites are selected. SR and hnRNP proteins binding to the mRNA act to define the exons and signal to the spliceosome where the splicing reaction is to take place.

Viruses by nature carry very little nucleic acid and many viruses have evolved to make use of the cellular splicing machinery in order to encode for multiple proteins from small amounts of genetic material. HPV is one such virus. Human papillomaviruses have a complex life cycle that involves a tight regulation of expression of the viral genes. Although papillomaviruses have eight genes they encode a large number of alternatively spliced RNAs which maximises their

coding capacity. For example, HPV16 encodes at least 13 late RNAs through extensive alternative splicing (Milligan et al., 2007).

2.2.1 SR Proteins

Serine/ Arginine rich (SR) proteins are the cellular splicing factors that are responsible for the identification of splice sites and recruitment of the components of the spliceosome (Wu & Maniatis, 1993, Staknis & Reed, 1994, Shen et al., 2004, Cho et al., 2011). For this reason they are considered positive splicing regulators. There are nine classical SR proteins that have a high degree of homology meaning there is some functional redundancy between the family members and are shown in Figure 1.10 (Shepard & Hertel, 2009). The family members have recently been renamed to have a standard nomenclature (Manley & Krainer, 2010). They are named SR proteins because they have a large serine/arginine rich domain. Each protein has at least one RNA recognition domain (RRM) as well as a serine/arginine rich domain. The RRM is involved in recognising and binding to specific splicing enhancer sequences in the target RNA (Long & Caceres, 2009). The RS domain is subject to phosphorylation and is involved in other protein-protein and protein-RNA interactions (Long & Caceres, 2009). Not only are SR proteins involved in splicing they are also involved in other RNA processing events in cells, such as mRNA export from the nucleus to the cytoplasm, mRNA stability and nonsense mediated decay (Huang & Steitz, 2001, Lemaire et al., 2002, Zhang & Krainer, 2004).

A lot of work has been carried out into understanding the functions of SR proteins as they seem to be involved in multiple stages of RNA processing. It is

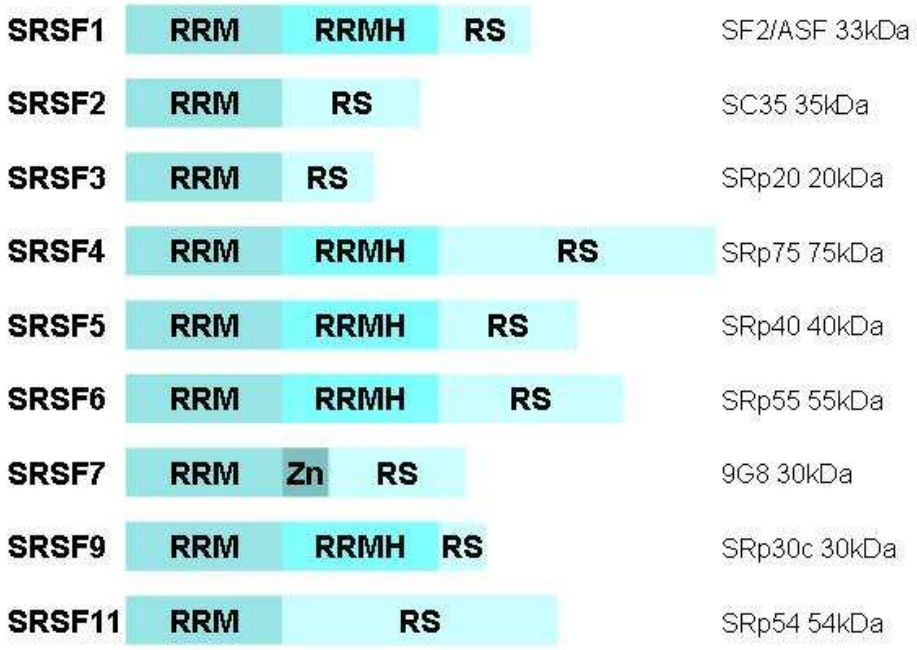


Figure 1.10 Schematic diagram of the 9 classical SR proteins. The protein structural domains of the 9 classical SR proteins are outlined in the diagram. Every SR protein contains at least one RNA Recognition Motif (RRM) and an RS domain (serine/arginine rich domain). Zn denotes a zinc motif and the RRMH domain denotes an RRM homology domain.

still unclear whether one SR protein family member is responsible for all of the SR mediated processing steps for a single RNA or whether multiple SR protein family members interact on an individual RNA. As SR proteins recognise specific sequences on the RNA at first it would seem likely that only one SR protein type is involved, however, due to the high redundancy in SR protein functions and a lack of true consensus binding sequences, it seems more likely that multiple SR proteins have the capability to be involved in the processing of individual RNAs. This would also explain why depletion of a single SR is not fatal to a normal cell if the SR proteins were in fact able to replace one another.

As previously mentioned SR proteins are involved in a number of RNA processing events as well as splicing. Certain SR proteins have been shown to shuttle continuously between the nucleus and the cytoplasm, e.g. SRSF3 and SRSF7, while other SR proteins, mainly SRSF2, have been shown to be nucleus confined (Caceres et al., 1998). This shuttling suggests that SR proteins may in fact be involved in transporting mRNAs from the nucleus to the cytoplasm for translation (Sanford et al., 2004, Michlewski et al., 2008). SRSF3 and SRSF7 were found to require a 22 nucleotide sequence on cargo RNAs and were found to work synergistically to export mRNA from the nucleus to the cytoplasm (Huang & Steitz, 2001). This ability of the SR proteins to bind and aid export of mRNAs may be of vital importance to those mRNAs that do not contain introns and therefore do not need to be spliced (Huang & Steitz, 2001). Dephosphorylation of the SR proteins was shown to be important for SRSF7 export functions in the nucleus and phosphorylation had to occur before the protein could re-enter the nucleus (Huang et al., 2004). In addition to this role in mRNA export SRSF1 has been shown to be associated with 80S ribosome after sucrose gradient fractionation suggesting a direct role in translation initiation (Sanford et al.,

2004). This translation enhancing role was demonstrated for both intron containing and intron lacking mRNAs, indicating that the translational effect is completely independent from the splicing activity of the SR protein (Sanford et al., 2004). SR proteins, e.g. SRSF2, have also been demonstrated to be involved in transcriptional elongation of certain mRNAs (Lin et al., 2008). Depletion of SRSF2 resulted in reduced levels of freshly formed polyadenylated mRNA due to a decrease in phosphorylation of RNA Polymerase II Serine 2 and therefore the stalling of transcription on actively transcribing genes (Lin et al., 2008). SR proteins also have a role in mRNA stability as some SR proteins can target mRNAs containing premature stop codons to the nonsense mediated decay pathway for degradation (Zhang & Krainer, 2004). In addition SRSF1 has been demonstrated to control the stability of the chicken PKCI-1-r mRNA (Lemaire et al., 2002). Binding of SRSF1 to a purine-rich sequence within the UTR of the PKCI-1-r mRNA, results in reduced stability of the RNA and its subsequent degradation (Lemaire et al., 2002). Thus SR proteins have multiple roles in RNA processing and can affect almost all stages from transcription to translation.

2.3 Polyadenylation and Cleavage

Once the mRNA has been transcribed and spliced the 3' end of the message must be polyadenylated. Polyadenylation again helps with the stability of the mRNAs. In eukaryotic RNAs the poly (A) tail can be between 200 and 300 nucleotides in length (Sheets & Wickens, 1989). Once transcribed the poly (A) signal (AAUAAA) on the RNA is then recognised and the RNA cleaved 10-30 nucleotides downstream of the signal allowing the RNA to dissociate with the polymerase,

and adenine residues added to the tail. Cleavage and polyadenylation involve a number of factors including cleavage/polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factors I and II (CF), poly (A) polymerase (PAP), symplekin and poly (A) binding protein nuclear 1 (PABN1) (for review see Zhang et al., 2010). CPSF recognises the RNA exiting the polymerase and recruits PAP to bind the RNA. CstF recognises the GU rich region of the RNA downstream from the poly (A) signal and PABN1 can bind to the poly (A) tail stimulating PAP to continue to add adenines to the 3' end of the RNA (Kuhn et al., 2009). The number of A residues added to the tail is variable and although at first may seem random, there is some level of control over this as there are de-adenylation enzymes which act to regulate the number of A residues on a particular RNA. A diagram of polyadenylation is shown in Figure 1.11.

2.4 Export

The final highly regulated step in the nuclear RNA processing pathway is mRNA export. Most mammalian mRNAs use TAP/p15 as their export adaptor (Cullen, 2003) although there are other export proteins. TAP and its cofactor p15 form heterodimers, which enhances the shuttling and nuclear pore binding capabilities of TAP (Wiegand et al., 2002). These export adaptors do not specifically recognise the RNA itself but proteins bound to the RNA, such as ALY/REF which is loaded onto the RNA during splicing (Zhou et al., 2000). ALY/REF has recently been shown to be recruited to the 5' end of the spliced mRNA in a cap- and splicing dependent process (Cheng et al., 2006). ALY/REF interacts directly with the cap-binding complex which the authors suggested

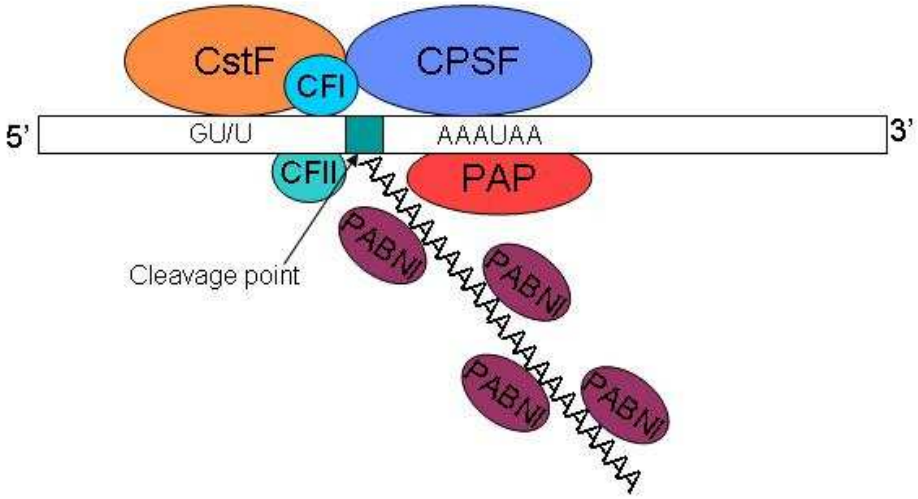


Figure 1.11 Schematic diagram of RNA polyadenylation. Cleavage stimulatory factor (CstF) (orange) binds to the GU rich region and cleavage/polyadenylation specificity factor (CPSF) (blue) binds to the polyadenylation signal of the extending RNA molecule and allows dissociation of the RNA from the polymerase. CPSF then recruits poly (A) polymerase (PAP) (red). Poly (A) binding nuclear protein 1 (PABN1) (purple) binds to the poly (A) tail and stimulates PAP to continue adding Adenine residues to the tail.

could explain why mRNA complexes are exported in a 5' to 3' direction (Cheng et al., 2006). ALY/REF is also a shuttling protein which enhances the shuttling of mRNAs from the nucleus to the cytoplasm (Zhou et al., 2000). SR proteins can also act to recruit the export adaptor TAP after losing some phosphorylation after splicing has occurred (Huang et al., 2003). Finally, it has also been shown that UAP56, which is involved in spliceosome assembly, recruits ALY/REF to the mRNA during splicing (Luo et al., 2001). All of these steps explain how spliced mRNAs can be exported out of the nucleus; however how the small numbers of unspliced or intronless mRNAs are exported from the nucleus remains unclear. When TAP/p15 is bound the mRNA is recruited to the nuclear pore complex, where TAP can bind the pore and is exported out of the nucleus into the cytoplasm (Cullen, 2003). Nuclear pore complexes (NPCs) mediate all communication between the cytoplasm and nucleus.

2.5 SR Proteins and Cancer

The prototypical SR protein SRSF1 (SF2/ASF) has recently been identified as being an oncoprotein (Karni et al., 2007). SRSF1 was shown to be upregulated in breast cancers due to chromosomal amplification of the SRSF1 gene found on chromosome 17q23 (Karni et al., 2007). Karni et al. (2007) showed that overexpressing SRSF1, but not SRSF2, in mouse fibroblast 3T3 cells was sufficient to transform the cells allowing anchorage independent growth in soft agar (Karni et al., 2007). When these SRSF1 overexpressing cells were injected into nude mice they formed large tumours, while cells overexpressing other SR proteins, SRSF2 and SRSF6, resulted in fewer and smaller colonies (Karni et al., 2007).

SRSF1 overexpression enhanced cellular proliferation and protected the cells from apoptosis. The expression of a number of proto-oncogenes and tumour suppressors was found to be altered upon overexpression of SRSF1 (Karni et al., 2007). SRSF1 overexpression resulted in the expression of a tumourigenic isoform of S6K1 (Karni et al., 2007). S6K1 is involved in the mTOR pathway and is involved in regulation of cell growth and apoptosis. When overexpressed, SRSF1 and SRSF2 resulted in an increase in isoform-2 production and a decrease in isoform-1 expression (Karni et al., 2007). Normally isoform-1 is the predominant isoform in the cell. Overexpression of SRSF6 did not alter the expression pattern of the S6K1 isoforms (Karni et al., 2007). When the isoforms were studied in order to test if they had any oncogenic functions, by overexpressing mouse isoforms in NIH 3T3 cells, overexpression of S6K1 isoform-2 resulted in colony formation in soft agar, whereas overexpression of isoform-1 did not allow for colony formation (Karni et al., 2007). However the colonies formed by overexpression of isoform-2 were not as numerous as the colonies formed by overexpressing SRSF1 alone, therefore other SRSF1 targets must be involved in the transformation process (Karni et al., 2007). Karni et al. (2007) showed that reducing the overexpression of SRSF1 back to relatively normal levels was enough to reverse the transformed state of the cells (Karni et al., 2007). They therefore concluded that SRSF1 could in fact be considered to be a proto-oncogene. However due to the high degree of homology between different SR proteins it is likely that SRSF1 is not unique amongst the SR proteins as possibly being oncogenic, SRSF2 and SRSR6 were also found to have some similar effects as SRSF1 during the study (Karni et al., 2007). Similar studies have been carried out on SRSF3 showing similar results. He et al. (2011) showed that by knocking down SRSF3 in ovarian cancer cells they could reduce cell

growth, inhibit anchorage independent growth and induce apoptosis in these transformed cells (He et al., 2011).

3 Aims

There are still some outstanding questions that have yet to be answered with regards to the current knowledge about HPV16 E6/E7 oncoprotein mRNA splicing. My hypothesis is that alternative splicing of HPV16 E6/E7 mRNAs is likely to be altered in response to epithelial differentiation status or transformation state of the HPV16-positive transformed cells. It is also likely that these isoforms carry out different viral functions, whether at the protein level or the RNA level. As host factors are required for cellular and viral processing it is probable that one or more of the splicing regulatory proteins that have been previously shown to be oncogenic in cancers and upregulated upon HPV16 virus infection is responsible for any changes in virus mRNA expression detected.

The first aim of this PhD was to investigate the alternative splicing patterns of HPV16 E6/E7 RNAs in the virus life cycle and during tumourigenesis. As shown previously, four isoforms of E6/E7 RNAs can be expressed (Tang et al., 2006). RT-PCR studies were performed in order to determine whether splicing of E6/E7 is altered upon differentiation of HPV16 infected epithelial cells or during HPV16 induced transformation of the epithelial cells using the W12 model of cervical tumour progression.

The functions of these E6/E7 mRNA isoforms were investigated in order to determine whether the alterations detected during the transformation or differentiation of the epithelial cells was related to any function. Mammalian expression constructs were generated expressing each of the individual E6/E7 isoforms. Functions of the different isoforms were assessed by expressing individual isoforms and carrying out growth curve studies and colony formation assays.

As alternative splicing is regulated by host cell splicing regulatory proteins the second aim for this PhD was to determine SR protein levels in transformed and virally infected cervical epithelial cells. Therefore the expression levels of a subset of SR proteins were investigated during W12 model cell line transformation and patient tissue. The levels of SR proteins in W12 model cervical epithelial cell lines were determined by western blotting and in patient tissues by immunohistochemistry.

To reveal any function for specific SR proteins in the HPV16-positive transformed cell lines, cells were siRNA treated in order to determine the SR proteins responsible for E6/E7 RNA processing. Thus RT-PCR experiments were performed and the SR protein responsible for E6/E7 mRNA processing during cervical epithelial cell transformation and HPV16 life cycle identified. Furthermore, HPV16 LCR transcription assays and RNA stability assays were preformed to define the level of control the SR protein exerts over E6/E7 mRNA processing.

Finally, as previous publications have shown that overexpression of some SR proteins have the ability to transform cells the final aim for this PhD is to determine affects of SR protein knockdown on transformed cervical epithelial

cells (Karni et al., 2007, He et al., 2011). Growth curve experiments, colony formation assays and Annexin V apoptosis assays were all performed.

2 Materials and Methods

2.1.1 Enzymes

All enzymes were purchased from Invitrogen (Life Technologies) unless otherwise stated.

2.1.2 Primers

All primers used in this thesis were purchased from Integrated DNA Technologies with the exception of the E6/E7 PCR primers and GAPDH primers which were purchased from Sigma-Genosys. Probe and primer sets were purchased from Eurogentec Ltd and predesigned and validated gene expression assays were purchased from Applied Biosystems.

2.1.3 siRNAs

All siRNAs used in this thesis were purchased from Dharmacon RNAi Technologies with the exception of the SRSF1 siRNA which was purchased from Sigma. All SR protein siRNAs are commercially available. SRSF1 and siGlo siRNAs were single molecule siRNAs and SRSF2 and SRSF3 were SMARTpool siRNAs.

2.1.4 Plasmids/Vector

pGEM®-T Easy (Promega): Taq DNA polymerase-amplified PCR products were cloned into this vector. The vector is supplied pre-cut with EcoRV and has 3' thymidine residues added to either end of the vector. The thymidine residues

complement the adenine residues added to the ends of amplified PCR products by Taq DNA polymerase. The inserted product is cloned into the ß-galactosidase gene which inactivates the enzyme, allowing detection of bacterial clones containing the ligated product by blue/white selection on Xgal/IPTG agar plates.

pEGFP-C1 (Clontech): A mammalian expression vector containing a CMV promoter and an SV40 polyadenylation site. PCR products cloned into the multiple cloning site will have an N-terminal GFP tag.

p3XFLAGCMV10 (Sigma): Mammalian expression vector containing a CMV promoter and an SV40 polyadenylation site. Inserts inserted into the multiple cloning site will have three FLAG epitopes added to the N-terminus.

p3XFLAGCMV10E6E7: Nucleotides 87 - 856 of HPV16 E6/E7 were cloned into the Hind III and Xba I sites of the p3XFLAGCMV10 vector.

pEGFP-C1-E6*I: Nucleotides 94-226^409-858 of HPV16 E6/E7 were cloned into the BglII and BamH1 sites of the pEGFP-C1 vector.

pEGFP-C1-E6*II: Nucleotides 94-226^526-858 of HPV16 E6/E7 were cloned into the BglII and BamH1 sites of the pEGFP-C1 vector.

pEGFP-C1-E6*X: Nucleotides 94-226^742-858 of HPV16 E6/E7 were cloned into the BglII and BamH1 sites of the pEGFP-C1 vector.

pGL3Control (Promega): Mammalian expression vector which contains an SV40 promoter and late polyadenylation site. The vector contains firefly luciferase which can be used to monitor expression of cloned products.

pGL3Basic (Promega): Mammalian expression vector containing an SV40 late polyadenylation site but no promoter.

pGL3HPV16LCR: Nucleotides 7101 – 137 of the HPV16 LCR were cloned into the pGL3-basic vector.

E2 Trans: HPV16 E2 encoding mammalian expression construct was a gift of Prof. Peter Howley.

GFPHPV6E2: GFP tagged HPV6 E2 mammalian expression construct was a gift of Dr Kevin Gaston.

GFPHPV11E2: GFP tagged HPV11 E2 mammalian expression construct was a gift of Dr Kevin Gaston.

GFPHPV16E2: GFP tagged HPV16 E2 mammalian expression construct was a gift of Dr Kevin Gaston.

Primer	Sequence 5’ to 3’	Use
E6 PrimerA F	GAGAACTGCAATGTTTCAGGACCC	PCR
E7 PrimerX R	GAACAGATGGGGCACACAATTCC	PCR
GAPDH F	TCCACCACCCTGTTGCTGTA	PCR
GAPDH R	ACCACAGTCCATGCCATCAC	PCR
E6 PrimerABglII F	AAAGGGAAAAGATCTGAGAACTGCAATGTTTCAGGACCC	Cloning
E7 PrimerXBamH1 R	GAACAGATGGGGCACACAATTCCATTGGATCCAAAGGGAAA	Cloning
T7	TAATACGACTCACTATAGG	Sequencing
SP6	GATTTAGGTGACACTATAG	Sequencing
pEGFP-C1	CATGGTCCTGCTGGAGTTCGTG	Sequencing

Table 2: Sequences of all primers used in this thesis and their applications

siRNA	Sequence	Final Concentration	Company Catalogue Number
HPV16 E6	AGU UAC CAC AGU UAU GCA C	40nM	Custom
siGlo RISC-free control	N/A	17nM	Dharmacon D-001600-01-20
SRSF1	CAU CUA CGU GGG UAA CUU A	20nM	Sigma SASI_Hs02_00313260
SRSF2 Smartpool	GCA CGA AGG UCC AAG UCC A	10nM	Dharmacon M-019711-00-0005
	CCA GUG UCC AAG AGG GAA U		
	CUC CCG AUG UGG AGG GUA U		
	GCG AGC UGC GGG UGC AAA U		
SRSF3 Smartpool	GAG UGG AAC UGU CGA AUG G	10nM	Dharmacon M-030081-00-0005
	GGA CUG UAA GGU UUA UGU A		
	CGA AGU GUG UGG GUU GCU A		
	CGA GAU GAU UAU CGU AGG A		

Table 3: Sequences and catalogue numbers of all siRNAs used throughout this project.

2.1.5 Antibodies

The antibodies used in western blotting and flow cytometry experiments are listed in the table along with the dilution they were used at and their source.

Antibody	Species	Dilution	Source	Catalogue Number
GAPDH clone 6CS	Mouse	1:1000	Biodesign International	H86504M
GFP 2555	Rabbit	1:500	Cell Signalling	2555S
γ-tubulin GTU88	Mouse	1:2000	Abcam	ab11316
SRSF1 clone 96	Mouse	1:1000	Zymed	32-4500
SRSF3 clone 7B4	Mouse	1:250	Zymed	33-4200
SRSF7 clone 98	Mouse	1:10	Dr James Stevenin	
Phosphorylated SR Proteins Mab104	Mouse	neat	ATTC hybridoma supernatant	
p53 clone DO-7	Mouse	1:500	BD Pharmingen	554294
pRb clone 4H1	Mouse	1:2000	Cell Signalling Technology	9309
HPV16 E2 TVG261	Rabbit	1:5000	Prof Iain Morgan	

Table 4: A table listing all antibodies, their source and dilutions used in this thesis.

2.1.6 Bacterial Culture

The DH5α strain of *Escherichia coli* was used to maintain all plasmid DNA. Cultures were grown in L-broth or on plates containing L-broth and 1.5% Bacto-agar. 100μg/ml ampicillin or kanamycin was added as appropriate.

2.1.7 Cell Lines

J2 3T3: A fibroblast cell line isolated from the J2 clone of random bred Swiss mouse.

W12E: An epithelial cell line containing around 100 episomal copies of the HPV16 genome (Stanley et al., 1989, Jeon et al., 1995).

W12G: An epithelial cell line containing 30 integrated copies of the HPV16 genome (Stanley et al., 1989, Jeon et al., 1995).

W12GPX: A transformed epithelial cell line containing integrated copies of the HPV16 genome. The cell line is feeder layer independent but mitogen dependent. W12GPX cells were derived from the W12G cell line (Aasen et al., 2003).

W12GPXY: A transformed epithelial cell line containing integrated copies of the HPV16 genome. The cell line is both feeder layer and mitogen independent. W12GPXY cells were derived from the W12GPX cell line (Aasen et al., 2003).

C33a: A transformed epithelial cell line negative for HPV DNA and RNA. Both p53 and pRb are present but are elevated in levels and of molecular mass respectively.

CaSki: A transformed epithelial cell line containing over 500 integrated HPV16 genomes.

SiHa: A transformed epithelial cell line containing 1-2 integrated HPV16 genomes.

HEK 293-T: Human embryonic kidney cell line stably transfected with the SV40 large T antigen.

2.1.8 Common Reagents, chemicals and solutions

30% Acrylamide/bisacrylamide mix	29% (w/v) acrylamide, 1% (w/v) N,N'-methylene bisacrylamide
10X PBS	1.37M NaCl, 26mM KCl, 100mM Na_2HPO_4, 18mM KH_2PO_4, pH7.4
1X PBS-Tween (PBS-T)	137mM NaCl, 2.6mM KCl, 10mM Na_2HPO_4, 18mM KH_2PO_4, 0.1% (v/v) Tween 20
10X TBE	0.89M Tris, 0.89M Boric Acid, 25mM EDTA
TE buffer	10mM Tris-HCl, 1mM EDTA, pH 8.0

DNA loading dye	50% glycerol, 1% SDS, 10mM Tris-HCl pH 8.2, Bromophenol blue
SDS protein loading buffer	125mM Tris pH 6.8, 4% SDS, 20% glycerol, Bromophenol blue, 5% ß-Mercaptoethanol

2.2 Methods

2.2.1 DNA manipulation and cloning

2.2.1.1 Bacterial transformation

Plasmid stocks and cloned vectors were maintained and amplified in DH5α *E. coli* (Invitrogen #18265-017). 10-50ng of plasmid or 5µl of ligated product was added to 50µl chemically competent cells and incubated on ice for 30 minutes. The reactions were heat shocked at 42°C for 45 seconds followed by a 5 minute incubation on ice. 450µl L-Broth was added to the cells and incubated at 37°C for 1 hour with shaking at 225rpm. 200µl culture was plated onto an L-agar plate with appropriate antibiotics and incubated at 37°C overnight.

2.2.1.2 Liquid Cultures

Small-scale bacterial cultures were prepared by inoculating 5ml of L-broth containing 100µg/ml of the appropriate antibiotic with a single colony of

bacteria. The cultures were incubated overnight at 37°C with shaking at 225rpm. In order to make a large culture 100ml L-broth containing antibiotics was inoculated with 100µl of a small culture.

2.2.1.3 Plasmid DNA extraction

Plasmid DNA was routinely extracted from small overnight bacterial cultures using the PureLink Quick Plasmid Miniprep kit (Invitrogen #K210010). The extractions were carried out according to the manufacturer's protocol. Briefly, 1.5ml of an overnight culture was pelleted by centrifuging at 13 200 rpm for 2 minutes in an eppendorf 5415 R bench top centrifuge and the media removed. The pellet was resuspended in 250µl resuspension buffer containing RNase A. 250µl lysis buffer, containing SDS to denature proteins, was added to the reaction and incubated at room temperature for 5 minutes before 850µl precipitation buffer was added. The reaction was added to a spin column and spun at 13 200 rpm for 10 minutes to remove cellular debris. The column was washed twice in 700µl wash buffer before the DNA was eluted in 75µl TE buffer into a new collection tube.

For transient transfections large bacterial cultures were inoculated and the PureLink HiPure Plasmid DNA Purification Maxiprep kit (Invitrogen # K210004) was used as the Miniprep kit can leave residual bacterial endotoxins in the DNA prep. The extractions were carried out according to the manufacturer's protocol. The initial steps are the same as described above only in larger volumes. For this procedure all centrifugation steps were carried out using a Beckman Coulter Avanti J-E centrifuge (JA-14 rotor). After eluting the DNA in 15

ml of elution buffer 10.5ml isopropanol was added to concentrate the DNA and remove salts. 5ml 70% ethanol was added to remove any remaining salts before the pellet was air dried and resuspended in 500µl TE buffer.

2.2.1.4 Restriction enzyme digestion

Plasmid or PCR-amplified DNA was digested using approximately 2U of restriction enzyme per µg of DNA. The reaction was carried out in the appropriate restriction buffer for 1-3 hours at 37°C.

2.2.1.5 Alkaline Phosphatase Treatment

In order to remove phosphates from the cut ends of digested plasmids CIP treatment was performed. This was especially important during cloning as BglII and BamHI were used to clone E6 isoforms. BglII and BamHI-cut vectors can religate without the addition of an insert. Purified cut plasmid was CIP digested in a volume of 50 µl. 5µl of 10X reaction buffer (New England Biolabs) and 1µl (10U) CIP enzyme (New England Biolabs) were added to 30µl (10-15µg) plasmid and the volume made up to 50µl. The reaction was incubated at 37°C for 1 hour.

2.2.1.6 Phenol:Chloroform extraction and ethanol precipitation

In order to remove excess salts and impurities from extracted RNA or DNA, a phenol:chloroform extraction was performed. An equal volume of phenol:chloroform was added to the nucleic acid and mixed thoroughly. Phenol:choloform was supplied at pH 4.3 for RNA extractions or pH 7.9 for DNA extractions. The reactions were centrifuged at 13 200 rpm in a bench top eppendorf 5415 R centrifuge for 5 minutes at room temperature. The upper aqueous phase was harvested to a clean eppendorf tube and one half volume of chloroform added and mixed thoroughly. Again the samples were centrifuged at 13200 rpm at room temperature for 2 minutes. The upper aqueous phase was harvested to a clean eppendorf tube. One tenth of the volume of 3M sodium acetate pH 5.2 was added followed by 2.5 volumes of 100% ethanol. The reactions were mixed thoroughly and stored at -20°C overnight or on ice for 30 minutes.

To harvest the nucleic acid, the samples were centrifuged at 13200 rpm for 10 minutes at room temperature in an RNase and DNase free 1.5 ml screw-capped eppendorf tube (Starstedt) and the supernatant removed. 5 volumes of 70% ethanol was added and centrifuged at 13200 rpm at room temperature for 10 minutes. The supernatant was removed and the nucleic acid pellet drained and dried followed by resuspension in either TE buffer (DNA) or DEPC treated water (RNA).

2.2.1.7 DNA ligation

PCR products were cloned into pGEM-T Easy vector (Promega) following the manufacturer's instructions. Briefly, 1µl (50 ng/µl) of pre-cut vector was added to 5µl of 2X rapid ligation buffer. Up to 3µl (~25ng) of purified PCR product was added followed by 1 µl (3U/µl) of T4 ligase. The reactions were incubated at 4°C overnight.

Digested inserts were ligated into the appropriately digested vector using T4 DNA ligase (10 000U/ml) (Invitrogen). The reactions were generally carried out in a 10µl volume with 1X ligation buffer (Invitrogen) and incubated for at least 2-3 hours at room temperature. A ratio of 3:1 insert: vector was typically used.

2.2.1.8 Agarose gel electrophoresis

DNA of greater than 1Kb in size was resolved in 0.7-1% (w/v) agarose gels (Roche) in 1XTBE. Agarose gels were electrophorised in 1XTBE buffer at 80V. DNA was stained in 0.5µg/ml ethidium bromide for 20 minutes before being visualised under UV. In order to gauge the relative size of the DNA fragments 1Kb DNA ladder (Invitrogen) was included in the gel.

2.2.1.8.1 Agarose gel DNA extraction

Extraction of DNA from agarose gels was carried out using the Qiagen Gel Extraction kit as described in the manufacturer's instructions. Briefly, the agarose slice containing the DNA was weighed and 5 volumes of buffer QG was

added and incubated at 50°C for up to 10 minutes or until the agarose had dissolved. 1 volume of isopropanol was added and transferred to a spin column. The samples were centrifuged at 13200rpm for 1 minute and the flow through discarded. The column was washed in wash buffer twice before the DNA was eluted into 30µl EB buffer and stored at -20°C.

2.2.1.9 Acrylamide gel electrophoresis

For DNA fragments less than 1kb, DNA was resolved on a 6% acrylamide gel. Acrylamide gels were composed of 3 ml of 30% acrylamide/bisacrylamide mix (Biorad), 1.5ml 10X TBE, 10.5ml dH_2O, 100µl of 10% APS and 12µl TEMED (Sigma). Gels were run in 1X TBE at 100V and the DNA stained in 0.5µg/ml ethidium bromide for 20 minutes before visualisation under UV using a Syngene Bio Imaging GeneFlash machine. 1Kb DNA ladder (Invitrogen) was included on the gel to calculate the size of the DNA fragment.

2.2.1.9.1 Acrylamide gel DNA extraction

To extract DNA from acrylamide gels, gel slices were incubated with 400µl 0.5M NaCl, 1mM EDTA overnight at 4°C. The acrylamide was spun out by centrifuging at 13200 rpm for 2 minutes in an eppendorf 5415 R bench top centrifuge and the DNA was precipitated by addition of one tenth volume of 3M sodium acetate pH 5.2 and 2.5 volumes 100% ethanol. After incubation at -20°C overnight the DNA was harvested by centrifugation at 13200 rpm for 10 minutes at room temperature and the supernatant discarded. 70% ethanol was added and centrifuged at 13200 rpm for 10 minutes at room temperature and the

supernatant discarded. The pellet was drained and dried before being resuspended in DEPC treated H_2O or TE buffer.

2.2.1.10 Nucleic acid quantification

The concentration of all extracted nucleic acids was determined using an Eppendorf BioPhotometer spectrophotometer. An A_{260} reading of 1 is equivalent to 40µg/ml of RNA and 50µg/ml double stranded DNA. The purity of the nucleic acid was determined using the A_{260}/A_{280} ratio. A reading of around 1.8-2.0 was considered to be reasonably pure.

2.2.2 Tissue Culture

2.2.2.1 Cell line growth and passaging

CaSki, SiHa, W12GPXY, C33a and 293-T cells were maintained in DMEM (Invitrogen) with the addition of 10% foetal bovine serum (Invitrogen). Cultures were passaged 2-3 times per week with splitting generally 1:5, 1:10 or 1:20 depending on the cell line. To passage the cells, the medium was removed and the cells washed in 1X sterile PBS and trypsinised using 1X trypsin-EDTA (Invitrogen) and incubated at 37°C until the cells detached. Cells were then resuspended into 10ml DMEM + 10% FBS and centrifuged at 1000 rpm for 5 minutes in a Heraeus Multifuge 3 S-R centrifuge and TTH-750 rotor. The pellet was resuspended in 10ml DMEM + 10% FBS and divided into 10cm plates at the

desired density. J2 3T3 cells were maintained in DMEM + 10% donor calf serum (Invitrogen).

W12E, W12G and W12GPX cells were maintained in DMEM + 10% FBS, 0.1nM Cholera Toxin and 0.4µg/ml Hyrocortisone. W12E and W12G cells require to be plated out on a feeder layer of J2 3T3 cells. In preparation for W12E or G cell growth, 3T3 cells were treated with 4µg/ml Mitomycin-C for 2-5 hours before harvesting the cells and counting. 3T3 cells were resuspended to $5x10^5$ cells in 4ml of culture media. $1x10^6$ W12 cells were added to the 3T3 cells and the cells plated out into five 10cm tissue culture plates, meaning $1X10^6$ 3T3 cells and $2X10^5$ W12 cells were added to each plate. After 24 hours, 0.5ng/ml epidermal growth factor was added to the cells. Cells were grown for 5 days for undifferentiated cells or 10 days for differentiated cells. While growing, the medium was changed and EGF added every two days to each plate until harvesting. To harvest the W12 cells, the 3T3 cells were first trypsinised and discarded before the W12 cells were harvested.

2.2.2.2 Cell stocks

To make stocks of cells, cells were harvested and counted before being resuspended to $1X10^6$ cells/ml in DMEM + 10% FBS + 10% Dimethylsulphoxide (DMSO) (Sigma). Cells were transferred into screw cap 2ml cryo-vials and stored at -80°C overnight before being transferred into liquid nitrogen for long term storage.

2.2.2.3 Transfection using Lipofectamine

2.2.2.3.1 Forward transfection using Lipofectamine RNAiMAX

In general, unless otherwise stated siRNA transfections were carried out as a forward transfection. 24 hours before transfection, $2x10^5$ of the appropriate cells were plated out in DMEM + 10% FBS without antibiotics. Transfections were most commonly carried out in 6 well plates. In a 1.5ml eppendorf, the required concentration of siRNA for a single well was diluted into 250µl Opti-mem serum free medium (Invitrogen) and in a separate 1.5ml eppendorf tube 5µl Lipofectamine RNAiMAX (Invitrogen) was diluted into 245µl Opti-mem medium. The complexes were combined and incubated for 20 minutes at room temperature before being added drop-wise to the cells. siRNA transfections were commonly carried out for 48 hours however for some experiments an incubation period of 24-72 hours was required. Transfection efficiency was monitored by the transfection of a fluorescently labelled control siRNA siGlo. siGlo is a fluorescently labelled siRNA which is excited at 557nm and is visible under the red channel. Transfection efficiency was monitored by counting the number of fluorescent cells in a field of vision and estimating the percentage efficiency.

2.2.2.3.2 Reverse transfection using Lipofectamine RNAiMAX

For some experiments a reverse transfection was more appropriate. The required concentrations of siRNA and 5µl Lipofectamine RNAiMAX were diluted into 500µl Opti-mem. The reagents were incubated at room temperature for 20 minutes before $2x10^5$ cells were added and plated out into a 6 well plate. 2.5ml

of medium without antibiotics was then added and incubated at 37°C for 24-72 hours. If siRNA knockdown and plasmid transfection were required to be carried out simultaneously this protocol was also followed with the addition of the plasmid to the initial dilution step before the addition of the cells.

2.2.2.3.3 Forward transfection using Lipofectamine

Unless otherwise stated plasmid transfections were carried out as a forward transfection. 24 hours before transfection, $2x10^5$ of the appropriate cells were plated out in DMEM + 10% FBS without antibiotics. Transfections were most commonly carried out in 6 well plates. In a 1.5ml eppendorf, the required concentration of plasmid was diluted into 100µl Opti-mem serum free medium (Invitrogen) and in a separate 1.5ml eppendorf tube 5µl Lipofectamine (Invitrogen) was diluted into 95µl Opti-mem medium. The complexes were combined and incubated for 40 minutes at room temperature. 800µl of serum and antibiotic free medium was added to the lipofectamine complexes before being added drop-wise to the cells. Before addition of the plasmid transfections the medium was removed from the cells and 800µl of serum and antibiotic free medium was added to the cells. The transfections were incubated at 37°C for 5-6 hours before the medium was replaced with DMEM + 10% FBS. Plasmid transfections were commonly carried out for 48 hours. Transfection efficiency was monitored by transfection with pMAXGFP (Lonza).

Cell type	Transfection efficiency
W12GPXY	80-90%
CaSki	70-80%
W12E/W12G	70-80%
C33a	70-80%
293-T	95-100%

Table 5: A table listing the transfection efficiencies obtained when transfecting the various cell lines with siGlo or GFP-plasmid in 293-T cells used during the course of the project.

2.2.2.4 Extraction of total cellular protein

Cells were washed twice in ice cold PBS after the removal of the culture medium. NP40 lysis buffer (0.5% NP40, 150mM NaCl, 50mM Tris HCl pH 8) with protease (Roche Diagnostics) and phosphatase (Roche Diagnostics) inhibitors (one tablet per 10ml of lysis buffer) was added to the cells and the cells scraped into the buffer on ice. The lysed cells were transferred to a 1.5ml eppendorf tube and incubated on ice for 30 minutes with periodic vortexing. The extracts were centrifuged at 13 200 rpm at 4°C for 10 minutes in an eppendorf 5415 R bench top centrifuge. The supernatant was transferred to a new eppendorf tube and stored at -80°C.

2.2.2.5 Extraction of total cellular RNA

After removal of the medium, the cells were washed in PBS and 1ml TRIzol (Invitrogen) added for every $1x10^6$ cells. The cells were scraped into TRIzol and transferred to a snap-cap Greiner RNase-free tube. The cells were incubated at room temperature for up to 10 minutes or stored at -20°C. One fifth volume of chloroform was added and mixed thoroughly and incubated at room temperature for 3 minutes. The RNA was centrifuged at 10 000 rpm at 4°C for 15 minutes in the Beckman Coulter Avanti J-E centrifuge (JA-14 rotor). The upper aqueous layer was carefully harvested into a new snap-capped tube. A half volume of isopropanol was added and incubated at room temperature for 10 minutes before centrifugation at 10 000 rpm for 10 minutes at 4°C. The supernatant was removed and the pellet washed in 70% ethanol. Again the RNA was centrifuged at 10 000 rpm for 10 minutes at 4°C. The ethanol was removed and the pellet drained and air dried before being resuspended in DEPC treated H_2O. RNA concentrations were determined as described in section 2.2.1.10.

2.2.2.6 Extraction of total cellular DNA

Cells were harvested by trypsinisation and counted. The cells were washed 3 times in ice cold PBS and centrifuged at 1000 rpm at room temperature for 5 minutes in a Heraeus Multifuge 3 S-R centrifuge and TTH-750 rotor. 300µl of lysis buffer (100mM NaCl, 10mM Tris-HCl pH 8.0, 25mM EDTA, 0.5% SDS, 0.1mg/ml Proteinase K) was added to 1-3 $X10^7$ cells. The extractions were incubated at 55°C overnight to completely lyse the cells. The DNA was then

phenol/chloroform extracted and ethanol precipitated as described in section 2.2.1.6.

2.2.2.7 Growth curve analysis

Cells were reverse transfected as described in section 2.2.2.3.2 with the appropriate siRNA and plated into 6 well culture plates. Cells were trypsinised and counted every 24 hours after plating for 72 hours and the total cell numbers calculated.

2.2.2.8 Colony formation assay

For colony survival assays, 1% agarose (Roche Diagnostics) in dH_2O, 0.7% agarose (Roche diagnostics) in dH_2O and 2X DMEM (Invitrogen) + 20% FBS (Invitrogen) were prepared. Cells were plated out 24 hours prior to transfection. In order to prepare the base layer, 1.25ml 1% agarose and 2.5ml 2X DMEM + 20% FBS for each 60mm plate was combined and plated before being stored at 4°C overnight after setting. Cells were plated out and a forward transfection carried out as described in sections 2.2.2.3.1 and 2.2.2.3.3. The transfections were incubated at 37°C for 24 hours. The top agarose layer was prepared by combining 1.25ml melted 0.7% agarose into 1.25ml 2X DMEM + 20% FBS for each 60mm plate. The cells were harvested and counted and 1 X 10^5 cells for each 60mm plate was added to the top layer. The top layer containing the transfected cells was plated out onto the base layer and incubated at 37°C for 12 – 14 days. The top layer was removed and 0.005% crystal violet staining solution was added and

incubated for 1 hour at room temperature. The stain was removed and the plates air dried before being photographed using a Sony DSC-P200 Digital camera.

2.2.2.9 Inverse invasion assay

To test the invasiveness of the cells, cells were plated out and transfected for 48 hours as described in section 2.2.2.3.1. Basement membrane matrix (VWR) was diluted 1:1.5 with ice cold serum-free DMEM and incubated on ice for 5 minutes. 60µl of diluted matrigel was pipetted carefully into support transwells (Fisher) added to a 24 well plate. The transwells containing the matrigel were incubated at 37°C for 1 hour to set. The transfected cells were harvested by trypsinisation and counted and resuspended to a concentration of $2X10^5$ cells/ml. After 1 hour the 24 well plate was removed from the incubator and turned upside down. The plate was carefully removed leaving the transwells standing upside down on the lid. $2x10^4$ cells or 100µl of resuspended cells was added to the bottom of the transwells and the plate carefully added to the lid. The 24 well plate was then incubated at 37°C for 2-3 hours upside down. 600µl of serum free medium was added to wells on a new 24 well plate. For every transwell, three wells containing medium were prepared. The new 24 well plate was incubated at 37°C until ready to use. After 2-3 hours the 24 well plate was turned the correct way up and the transwells washed twice in two of the wells containing the serum free medium. The transwell was placed into the remaining well containing serum-free medium and 150µl of DMEM + 10% FBS was added to the centre of the well creating a concentration gradient of serum. The transwells were incubated at 37°C for 3 days. After 3 days cells were stained in 4µM

Calcein-A. 4mM Calcein-A was diluted 1:1000 into serum-free DMEM and HEPES (100:1 DMEM:HEPES). 500µl of diluted Calcein-A was added to a 24 well plate. The transwells were removed from the medium and placed into the 24 well plate containing staining solution. 500µl of diluted Calcein-A was added into the transwell and the plate incubated at 37°C for 1 hour. The cells were then viewed under the fluorescence Leica DMIRE2 microscope. Z stack images were taken every 10µM to determine invasion potential.

2.2.2.10 Cell cycle analysis

Cells were plated out at a density of $1X10^6$ cells and were transfected as described in section 2.2.2.3.1. After 48 hours the cells were harvested by removing the medium washing in PBS twice. Cells were trypsinised and collected by centrifugation at 1000 rpm at 4°C for 10 minutes in a Heraeus Multifuge 3 S-R centrifuge and TTH-750 rotor. The cell pellet was washed in PBS and centrifuged under the same conditions as above another two times. The samples were then resuspended in 500µl of PBS and placed onto ice immediately. The cells were fixed by the addition of 4.5 ml of ice cold 70% ethanol added drop wise over a vortex at low speed. The cells were incubated on ice for 30 minutes, or if not required immediately, left at -20°C. For DNA analysis, the cells were stained with Propidium Iodide. The samples were centrifuged under the same conditions as before and the ethanol removed. The pellet was again washed twice in PBS as before. The cells were resuspended in 500µl of PBS and 10µg/µl of Propidium Iodide and 10µg/µl Ribonuclease A added. The staining was carried out for at least 3 hours at 4°C. The samples were

analysed using the EXPO32ADCXL4 Colour program on a Beckman Coulter Epics XL-MCL machine.

2.2.2.11 Transcription assay

293-T cells were transfected as described in section 2.2.2.3.3 and incubated at 37°C for 48 hours. Cells were washed twice in PBS before 300µl 1X Reporter Lysis Buffer (Promega) was added and incubated at room temperature for 15 minutes. Cells were scraped into the reporter lysis buffer and transferred to -80°C overnight. Following an overnight incubation, the samples were defrosted on ice and centrifuged at 13 200 rpm at 4°C for 30 minutes in an eppendorf 5415 R bench top centrifuge before the supernatant was transferred to a new eppendorf tube. 80µl of extract was added per well of a white 96 well plate (Fisher Scientific catalogue number: MPA-560-030D). Luciferase assay substrate (Promega) was dissolved into Luciferase assay buffer (Promega) and aliquoted and stored at -80°C. 1 ml of resuspended luciferase assay substrate was diluted into 3 ml dH_2O and the plate read by a Thermolabsystems Luminoskan Ascent plate reader. The plate reader adds 80µl diluted luciferase substrate to each well and measures the luciferase activity.

2.2.3 Protein Analysis

2.2.3.1 Protein Quantification

10µl of protein extract was added to 1ml diluted 1X Bradfords reagent (Biorad) in a plastic cuvette. BSA (Sigma) standards (0.1mg/ml, 0.2mg/ml, 0.4mg/ml, 1mg/ml and 2mg/ml) were also prepared and the absorbance at wavelength A_{595} was measured. The concentrations of the protein samples were determined after a standard curve of protein standards was produced and the equation of the straight line solved for each sample (y=mx+c).

2.2.3.2 SDS-PAGE

Equal concentrations of extracted proteins were added to 1X NuPAGE LDS protein loading dye (Invitrogen) and 1X NuPAGE sample reducing agent containing 500mM DTT (Invitrogen) before being boiled at 100°C for 5 minutes to denature any secondary structures that may be present in the protein sample. Samples were resolved on NuPAGE Novex 4-12% gradient Bis-Tris polyacrylamide gels (Invitrogen) in 1X MES running buffer (Invitrogen). Electrophoresis was carried out at 200V.

2.2.3.3 Western Blotting

After electrophoresis, protein gels were transferred onto nitrocellulose membranes using the Invitrogen iBlot system. The iBlot system is a dry blotting

method of protein transfer and takes 7 minutes to complete. Membranes were blocked in 5% non-fat Marvel powdered milk in PBS-0.1% Tween20 for 1-2 hours at room temperature. Membranes were washed briefly in PBS-T before being incubated at 4°C overnight with rotation with the desired primary antibody. The primary antibodies and their dilutions are listed in table 3. The following day the membranes were washed in PBS-T (1X 15 minute wash followed by 3X 5 minute washes) and incubated with the appropriate secondary antibody for 2 hours at room temperature. Secondary antibodies (GE Healthcare) were diluted 1:2000 in 2.5% non-fat Marvel in PBS-T. All secondary antibodies were horseradish peroxidase tagged. Membranes were washed in PBS-T (as above) before being detected by chemiluminesence by incubating with either ECL (Pierce) for 1 minute or with ECL (plus) (GE Healthcare) for 5 minute depending on the primary antibody. Proteins were detected using Kodak 18X24 mm Medical X-Ray film and the Xograph Compact X4 film processor.

2.2.3.4 TUNEL staining

Cells were plated out in a 4 well chamber slide ready for forward siRNA transfection as described in section 2.2.2.3.1. After 24-72 hours of transfection the medium was removed from cells and washed twice in 1ml PBS. TUNEL staining was carried out using the DeadEnd Fluorometric TUNEL system (Promega). The cells were then fixed with 1ml 4% formaldehyde for 25 minutes at 4°C. The formaldehyde was removed and the cells washed twice in PBS for 5 minutes at room temperature. The PBS was removed and the cells permeabilised in 1ml 0.2% TRITON X-100 in PBS at room temperature for 5 minutes. The cells were then washed twice in 1ml PBS for 5 minutes. A positive

control was set up by adding 100µl of RQ1 DNase-1 buffer (Promega) and incubating at room temperature for 5 minutes. The buffer was replaced with another 100 µl DNase-1 buffer containing 10U/ml RQ1 DNase-1 and incubated at room temperature for 10 minutes. The buffer was removed and the cells washed in deionised water 4 times. 100µl Equilibration buffer (Promega) was added and incubated at room temperature for 10 minutes. The staining solution comprising of 45µl Equilibration buffer, 5µl Nucleotide Mix (50 µM fluorescein-12-dUTP, 100 µM dATP, 10mM Tris-HCl (pH 7.6), 1mM EDTA) and 1µl rTdT enzyme per well was prepared and 50µl of the staining solution was added to each well and incubated in humidified chamber at 37°C for 1 hour. The reaction was stopped by filling the well with 2X SSC (Promega) and incubating at room temperature for 15 minutes. The cells were washed three times in 1ml PBS at room temperature for 5 minutes. The wells were air dried before 1 drop of Vectashield + DAPI was added and a glass coverslip placed on top. The Vectashield was allowed to set before the slide was viewed under the Leica DMIRB fluorescent microscope or stored overnight at 4°C in the dark.

2.2.3.5 Annexin V staining

Cells were transfected as described in section 2.2.2.3.1 and harvested by trypsinisation. Floating cells were also collected by centrifugation of the medium. The cells were washed in ice cold PBS and centrifuged at 1000 rpm at room temperature for 5 minutes. The cells were resuspended at a concentration of $1X10^6$ cells/ml in annexin binding buffer (10mM HEPES, 140mM NaCl, 2.5mM $CaCl_2$ pH 7.4). To 100µl of resuspended cells, 5µl of annexin V conjugate 488nm (Invitrogen) and 1µg/ml propidium iodide was added. The staining cells were

incubated at room temperature in the dark for 15 minutes before 400µl of annexin binding buffer was added and placed on ice. Samples were analysed on a BD Biosciences FACScalibur machine.

2.2.3.6 Senescence associated β-galactosidase staining

The senescence ß-gal staining kit (Cell Signalling Technology) was used to measure senescence in transfected cells. Cells were transfected as described in section 2.2.2.3.1 and incubated for 24 – 72 hours. The medium was removed from the cells and they were washed twice in 1ml PBS. The cells were fixed in 1ml 1X Fixative solution for 15 mins at room temperature. Meanwhile staining solution composed of 930µl staining solution, 10µl staining supplement A, 10µl staining supplement B and 50µl X-Gal for each well was prepared. The cells were washed twice in 2ml PBS before 1ml staining solution was added and incubated overnight at 37°C without 5% CO_2. The cells were visualised the next day under the Leica DMIL LED light microscope or the staining solution was removed and 1ml 70% glycerol was added and stored at 4°C.

2.2.4 RNA Analysis

2.2.4.1 RT-PCR

2.2.4.1.1 mRNA extraction

Following RNA extraction as described in section 2.2.2.5, mRNA was extracted from the total cellular RNA using the Qiagen Oligotex mRNA extraction kit as

described in the manufacturer's instructions. Briefly, 250µg of RNA in 250µl of DEPC-H_2O was added to 250µl of Buffer OBB and 15µl of Oligotex solution and mixed. The RNA was incubated at 70°C for 3 minutes to denature any secondary structure before incubating at 20°C for 10 minutes to allow the poly A tail of the mRNA to bind to the Oligotex particles. The RNA was centrifuged at 16110 g for 2 minutes in an eppendorf 5415 R bench top centrifuge and the supernatant removed. The Oligotex pellet was resuspended in 400µl Buffer OW2 and transferred to a spin column. The column was centrifuged at 16110 g for 1 minute and the flow through discarded. The column was transferred to a new 1.5ml eppendorf tube and 400µl Buffer OW2 applied to the column. The RNA was again centrifuged at 16110 g for 1 minute and the flow through discarded. The spin column was again transferred to a new 1.5ml eppendorf tube and 20µl of Buffer OEB preheated to 70°C was added and the mRNA resuspended. The mRNA was centrifuged at 16110 g for 1 minute before a second 20µl of hot Buffer OEB was added to maximise the mRNA yield. The mRNA was centrifuged again at 16110 g for 1 minute and the eluted mRNA stored at -20°C.

2.2.4.1.2 DNase-1 treatment for RT-PCR

After RNA extraction as described in section 2.2.2.5, any remaining DNA in the RNA extract was removed by DNase-1 digestion using the Promega RQ1 DNase-1 kit as described in the manufacturer's instructions. 1U of RQ1 was used per 1µg of RNA. Reaction buffer to a final concentration of 1X was added to the reaction and the digestion carried out for 1 hour at 37°C. 1µl of DNase-1 stop solution was added per µg of RNA and incubated at 65°C for 10 minutes to stop the reaction. The RNA was then phenol:chloroform extracted as described in section

2.2.1.6 to remove enzyme and excess salts from the sample that may inhibit downstream applications.

2.2.4.1.3 Reverse Transcription

After DNase-1 digestion, the RNA was reverse transcribed into cDNA using the Superscript III kit (Invitrogen) following the manufacturer's instructions. 1µl of 10mM dNTP mix and 1µl of Oligo dT_{20} (50µM) primer was added to 1µg of RNA and incubated at 65°C for 5 minutes to disrupt any secondary structures. The reactions were placed on ice and 2µl 10X RT buffer (200mM Tris-HCl (pH 8.4), 500mM KCl), 4µl 25mM $MgCl_2$, 2µl 0.1M DTT, 1µl 40U/µl RNase OUT and 1µl (1U/µl) of either Superscript III or DEPC-treated H_2O (for the RT- controls) were added. The reactions were incubated at 50°C for 50 minutes before being incubated at 85°C for 5 minutes to stop the reaction. The samples were placed on ice and 2U (1µl) RNase H was added and incubated at 37°C for 20 minutes to degrade the initial mRNA strands.

2.2.4.1.4 PCR

2µl of cDNA (~100µg) or 1µl (~0.5µg) of miniprep DNA was amplified in a PCR reaction using Taq DNA polymerase (Invitrogen). Reactions were carried out in 50µl volumes with 38.1µl DEPC treated H_2O, 5µl 10X PCR Buffer (200mM Tris-HCl (pH 8.4), 500mM KCl), 1.5µl 50mM $MgCl_2$, 1µl 10mM dNTPs, 1µl gene specific forward primer (10µM), 1µl gene specific reverse primer (10µM), 2µl cDNA and 0.4µl 5U/µl Taq DNA polymerase. The final concentration for the PCR reaction was 1X PCR Buffer, 200nM Primers, 200µM dNTPs, 1.5mM $MgCl_2$ and 2U Taq DNA

polymerase. PCRs were carried out on a Hybaid PCR Express machine with a general program consisting of:

95°C – 2 minutes

95°C – 30 seconds
55°C – 30 seconds
72°C – 1 minute
} 25-30 cycles

72°C – 5 minutes

55°C – 5 minutes

72°C – 10 minutes

4°C – hold

PCR products were resolved on acrylamide gels as described in section 2.2.1.9.1.

2.2.4.2 qRT-PCR

2.2.4.2.1 DNase-1 Treatment for qPCR

After RNA extraction as described in section 2.2.2.5, any remaining DNA in the RNA extract was removed by DNase-1 digestion using the Ambion TURBO DNase-1 kit as described in the manufacturer's instructions. RNA was diluted to 50ng/µl in a 30µl volume. 3µl 10X reaction buffer and 1µl of 2U/µl TURBO DNase was added to the reaction and the digest carried out for 30 minutes at 37°C. 3.5µl of Inactivation Reagent was added and incubated at room temperature for 5 minutes to stop the reaction. The RNA was then centrifuged at 13 200 rpm for

1.5 minutes in an eppendorf 5415 R bench top centrifuge and the supernatant recovered to a fresh tube. The centrifugation step was repeated and the recovered RNA stored at -20°C.

2.2.4.2.2 Reverse transcription

After DNase-1 digestion the RNA was reverse transcribed into cDNA using the Affinityscript qPCR cDNA synthesis kit (Stratagene) following the manufacturer's instructions. 10µl of 2X Mastermix, 2µl of Oligo dT_{20} primer, 1µl random primers (0.1µg/µl) and either 1µl Affinityscript RT/RNase Block enzyme was added to 6µl of DNase-1 treated RNA and incubated with the following program on the Hybaid PCR Express machine:

25°C – 5 minutes

42°C – 45 minutes

95°C – 5 minutes

4°C - hold

2.2.4.2.3 qPCR

After reverse transcription or DNA extraction, cDNA or DNA was amplified using an Applied Biosystems 7500 qPCR machine. For some targets, for example many human genes, probe and primer sets are commercially available and have been optimised and shown to have close to 100% efficiency. Therefore when possible these Taqman predesigned gene expression assays were ordered. However for other targets, for example viral genes, there are no commercially available assays, therefore these assays had to be custom designed and synthesised. The Stratagene Brilliant qPCR mastermix was used for qPCR reactions. Reactions

were prepared as follows: 12.5µl of 2X Mastermix, 2.25µl of 10µM forward primer, 2.25µl of 10µM reverse primer, 0.5µl of 5µM probe, 0.5µl of 15µM reference dye, 5µl of DEPC-water and 2µl cDNA or 100ng DNA. If a Taqman predesigned gene expression assay was being used then the reactions are prepared as follows: 12.5µl of 2X Mastermix, 1.25µl gene expression assay, 0.5µl of 15µM reference dye, 10.75µl of DEPC-water and 2µl (~15ng) cDNA or 100ng DNA. The amplification protocol consisted of 50°C for 2 minutes, 95°C for 10 minutes, 40 cycles at 95°C for 15 seconds followed by 60°C for 1 minute. To calculate the relative efficiencies of probe and primer sets, cDNA or plasmid standard curves were generated and the efficiency calculated from the slope of the line using the formula Efficiency = $10^{(-1/slope)}$-1. For probe and primer sets to be used in a relative quantification calculation the efficiencies must be within +/- 10% of each other. The predesigned assays were all guaranteed to be 100% efficient therefore no standard curve was performed. Relative quantification was not used with these assays. The relative efficiencies for my custom E6 isoform probe and primer sets were extremely varied out with +/- 10% of each other and could not be used for relative quantification.

2.2.4.3 DNA sequencing

Sequencing of PCR products or cloned vectors was carried out on an ABI PRISM 3130*XI* Genetic Analyzer machine. All sequencing reactions were carried out using the Applied Biosystems Big Dye Mix and 5X sequencing buffer. The DNA to be sequenced was amplified as follows: 2 µl Big Dye Mix, 4 µl 5X Sequencing Buffer, 2 µl 10 µM sequencing primer, 6 µl ddH2O and 6 µl (~3µg) of DNA from a

standard miniprep. The products were amplified using the PCR Express machine as follows

95°C - 10 sec
50°C - 5 sec
} 25 cycles
60°C - 4 min
4°C - hold

Products were purified using the EdgeBio Spin column purification method. The columns were centrifuged for 2 minutes at 3K in an eppendorf 5415 R bench top centrifuge to remove storage solution. The collection tube was discarded and the spin column inserted into a fresh collection tube. The PCR sequencing reaction was transferred onto the centre of the gel in the spin column and the column centrifuged for 2 minutes at 3K in an eppendorf 5415 R bench top centrifuge. The DNA was dried under vacuum for 15-20 minutes and the pellet resuspended in 25 µl of Hi-Di Formamide (Applied Biosystems).

3 Results 1

3.1 Introduction

A large body of scientific evidence has led to the conclusion that E6 and E7 are the major oncoproteins of the high risk HPVs. However, one problem with our current knowledge is that the high risk virus E6 and E7 mRNAs can be alternatively spliced yet it is not known which E6/E7 mRNA isoform(s) contribute to the transformed phenotype. Interestingly, low risk virus E6/E7 mRNAs appear not to be alternatively spliced suggesting that high risk HPV oncogenes have evolved to encode alternative protein isoforms that contribute to the virus replication cycle in some way but also may be involved in cancer progression. In particular, HPV16 E6 can be alternatively spliced with the production of at least four isoforms. The first reports of E6 being alternatively spliced came in 1986 when Smotkin and Wettstein identified two spliced isoforms of HPV16 E6 after RT-PCR analysis of RNA isolated from transformed cells and tumour biopsies (Smotkin & Wettstein, 1986). In this study they identified the largest spliced product, now called E6*I as being the predominant isoform in transformed cells. It was suggested that this spliced product in fact allows for efficient translation of E7 protein. The initiation codon for E7 is only two bases downstream from the termination codon of E6. Therefore the hypothesis is that there is not enough RNA between the two open reading frames to allow for efficient translation re-initiation. Splicing of the RNA and termination of the E6 isoform protein at an earlier stop codon allows more space between the E6 isoform termination codon and the E7 start codon. However, further studies have since

demonstrated that E7 is translated as efficiently from E6/E7 bicistronic mRNAs as from spliced E6 mRNAs (Stacey et al., 1995). Furthermore, the assumption from these earlier studies was that, as E6*I is the predominant isoform in cervical tumour cells, the isoform may have tumour promoting functions. However, a number of studies over many years have suggested that this is unlikely to be true. As will be discussed, E6*I appears to have opposing functions to E6 with regards to a number of cellular interacting partners.

In 2006, Tang et al. reported the possibility of four mRNA splice isoforms being produced from the open reading frames encoding E6 and E7 (Figure 3.1) (Tang et al., 2006). One splice donor site was located in the E6 open reading frame with two splice acceptor sites in the E6 open reading frame and one in the E7 open reading frame (Tang et al., 2006). The two splice acceptor sites in the E6 coding region were those already identified in previous studies: E6*I and E6*II. A third acceptor site was identified in the E7 region and the putative isoform was named E6*X. There have been comparatively few reported functions for E6*I compared to the number of reports of full length E6 functions and no reported functions for E6*II and E6*X.

The first report of any function for HPV16 E6*I came from Shirasawa et al. in 1994 where they showed that while E6 has a repressive effect on the P_{97} major early promoter, low levels of E6*I could in fact *trans*-activate the promoter, indicating that E6*I may have completely separate functions to E6 (Shirasawa et al., 1994). The next report of E6*I function came from studying the effects of E6

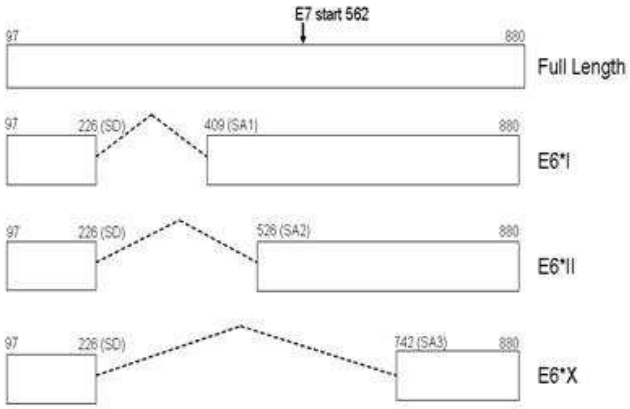


Figure 3.1 Diagram of HPV16 E6 isoforms. Schematic diagram of the E6/E7 isoforms with the splice donor (SD) and acceptor (SA) sites labelled and the nucleotide numbers given.

and E6*I on p53. As already stated, E6 targets p53 for proteasomal degradation to inhibit growth arrest and apoptosis in HPV-infected cells. Shalley et al. (1996) and Pim et al. (1997) studied HPV16 E6 and HPV18 E6 respectively and found that the spliced E6 product acts to protect p53 from the degradation induced by full length E6 proteins (Shally et al., 1996, Pim et al., 1997). Pim et al. (2007) also demonstrated that expression of E6*I in CaSki cervical cancer cells decreased the levels of cellular proliferation similar to expression of p53 in CaSki cells (Pim et al., 1997). This finding suggests that E6*I, at least for HPV18, is in fact acting to suppress the transformed phenotype of the cells.

E6 has been shown to have anti-apoptotic effects, therefore many of the studies into E6*I function have analysed levels of apoptosis. E6*I was shown to have an opposing effect from E6 full length on procaspase 8. Filippova et al. (2007) reported that the full length E6 protein could bind to procaspase 8 and accelerate its degradation thus preventing U2OS cells from entering apoptosis (Filippova et al., 2007). In contrast, although the smaller E6*I isoform could still bind procaspase 8, it stabilised it thus making the cells susceptible to apoptosis (Filippova et al., 2007). The authors explained this seemly paradoxical interaction of the two E6 isoforms with procaspase 8 by suggesting that the ratio of the two isoforms may be the important factor in determining procaspase 8 levels and therefore the cells' potential to enter apoptosis (Filippova et al., 2007). Furthermore, the procaspase 8 binding domains of E6 and E6*I have been identified and are not the same, perhaps explaining why two similar proteins can have such diverse effects (Filippova et al., 2007). This suggests that there could be a very delicate balance of E6 expression in HPV-infected cells. Similar results

have been demonstrated for FADD and TNF R1 where both the full length E6 and E6*I have opposing effects on stabilising or destabilising the molecules upon protein-protein interaction (Filippova et al., 2002, Filippova et al., 2004, Filippova et al., 2009). The targeting of these apoptosis effector molecules shows that E6 very tightly regulates the extrinsic apoptosis pathways. In these studies E6 and E6*I were shown to be able to form complexes which may also be important in controlling E6 action (Filippova et al., 2009). The levels of E6 in the cell were also reported to be important as low levels of E6 and E6*I make cells resistant to TNF induced apoptosis (Filippova et al., 2009). E6-mediated control over apoptosis pathways is extremely complex and it appears to require a very delicate balance of E6 levels and isoform ratios.

Pim et al. (1997) also demonstrated a similar scenario when studying the E6-mediated degradation of PDZ domain-containing proteins. They showed that E6*I could direct the degradation of the tumour suppressor hDlg as efficiently as E6 full length and that the two proteins appeared to cooperate to allow efficient degradation (Pim et al., 1997). Again the ratio between the two isoforms was reported to be important for regulating the function of E6 (Pim et al., 1997).

Finally, a recent report showed that splicing of E6 is regulated by the EGF pathway. Rosenberger et al. (2010) demonstrated the requirement for EGF for production of full length E6 transcripts and that a reduction in EGF levels altered the E6 splicing pattern into producing mainly E6*I transcripts (Rosenberger et al., 2010). This finding is very interesting as the authors suggested that perhaps

full length E6 is required initially to avoid apoptosis after infection (Rosenberger et al., 2010). There is a concentration gradient of EGF in the differentiating epithelium and the levels of EGF are highest in the basal epithelial layer. Rosenberger et al propose that as the infected cell moves up through the epithelium the requirement for full length E6 may change and E7 production, to keep the infected cell cycling, becomes more important. Therefore the EGF levels drop, E6*I transcripts becomes more predominant and translation of E7 is enhanced (Rosenberger et al., 2010). This hypothesis is still controversial because there is conflicting evidence that E6*I allows enhanced E7 production. However, this study provides the first suggestion of a switch between isoform production being important for the virus life cycle. It is also important to note that the E6*I transcript is the dominant isoform in transformed cells and the levels of E6*I have been reported to be indicative of severity of the clinical lesion (Kosel et al., 2007). This may also fit with our current knowledge of E6 splicing control: if lesions express higher levels of E6*I and if the original hypothesis of alternative splicing of E6 providing enhanced translation of E7 is true, and there is a proliferative advantage of expressing high E7 levels, this may explain why E6*I levels are high in lesions.

As already mentioned there have so far been no reported functions for E6*II and E6*X. In my opinion, however, due to the vast number of E6 protein-interacting partners already identified, I think it is entirely possible that one or more of the E6 isoforms may be involved in some of the functions attributed to the E6 protein. For example all isoforms are spliced from the same splice donor site therefore all proteins have the same N-terminus. Given the presence of the

same N-terminus in all E6 isoforms these could still carry out any functions mapped to this domain in the full length E6 protein, for example part of the first zinc finger domain is retained in the E6 isoforms however it will be unable to form the loop finger structure, therefore E6AP could still bind if the structure is not important for binding. All isoforms lack the entire C-terminus of E6 including the second zinc finger domain and the whole PDZ domain. Splicing in all isoforms occurs at splice donor 226 which is equivalent to amino acid 41 meaning that the first CXXC motif at amino acid 30 is present in all isoforms but the second at amino acid 63 is absent. This means the zinc finger structure cannot form; therefore it is likely that interactions dependent on this first zinc finger domain will also be lost. E6*I is a small protein and after splicing only two additional amino acids are added after the splice junction before meeting a frameshift termination codon. Therefore it is likely that only interactions attributed to full length E6 N-terminus (amino acids 1-41) will be able to be fulfilled by E6*I. However as E6*I still conatins the L1 and S1 domains (Figure 1.5) it could retain the capacity to bind full length E6 and block folding between the N-terminus and C-Terminus of E6. This could explain why E6*I can antagonise full length E6 function if there is a dominant negative effect of E6*I binding to full length E6. The same could be true of all isoforms as they contain the same portion of the N-terminus. This however does not explain any transformation properties of the isoforms and at present these properties cannot be determined or mapped to the structures of the isoforms. E6*II has an additional 22 amino acids added after the splice junction before meeting a stop codon, however there is a frame shift meaning the amino acids are not the same as in full length E6 protein, therefore the structure and the interactions in this region will be different. It is not possible to determine what the structure of E6*II may be as there is no crystal structure for the altered amino acids and the

presence of these different amino acids could in fact alter the folding of the N-terminus. E6*X is spliced into E7 therefore it is possible that functions attributed to the N-terminus of E6 and the C-terminus of E7 will be the same. Importantly E6*X is spliced in frame with the 3' end of the E7 open reading frame therefore these interactions may be the same as full length E7. For example E6*X could still interact with E2F6 which interacts with the extreme C- terminus of HPV16 E7 (McLaughlin-Drubin et al., 2008). As detailed by the crystal structure, the C-terminus of HPV45 E7 is a highly structured domain, whereas the N-terminus is not structured, therefore as E6*X is spliced in frame and expresses the C-terminus of E7 the function of E6*X could be the same as full length E7. Similar to the full length E7 protein it is also possible that E6*X could form dimmers due to the highly structured C-terminus.

Due to the lack of sensitive antibodies for the detection of endogenous E6 protein levels, it is not known if E6*II and E6*X mRNA isoforms can produce functional proteins, however even if no protein is produced from the RNA, the RNAs themselves could still have a function. For example, they could be controlling viral or cellular gene expression. It has recently been shown that a class of long non coding RNAs called lncRNAs can epigenetically regulate expression of target genes (Kaikkonen et al., 2011). LncRNAs have been proposed to recruit chromatin remodelling complexes to target genes with homology to the lncRNA and silence their expression. Other ncRNAs have been found to associate with the enhancer and promoter elements of target genes and alter their expression (Kaikkonen et al., 2011). It has also been demonstrated that ncRNAs can also interfere with post-transcriptional gene expression by inhibiting splicing by masking splice sites and preventing spliceosome binding (Kaikkonen et al., 2011). It is also possible that the original hypothesis proposed

by Sedman et al. (1991) is true; that E6 is spliced to allow for enhanced and efficient translation of E7. While this could be true for E6*I and E6*II, it seems unlikely for E6*X as the RNA is spliced from E6 into the middle of the E7 open reading frame producing a hybrid RNA which could therefore have a completely different function from either E6 or E7 alone.

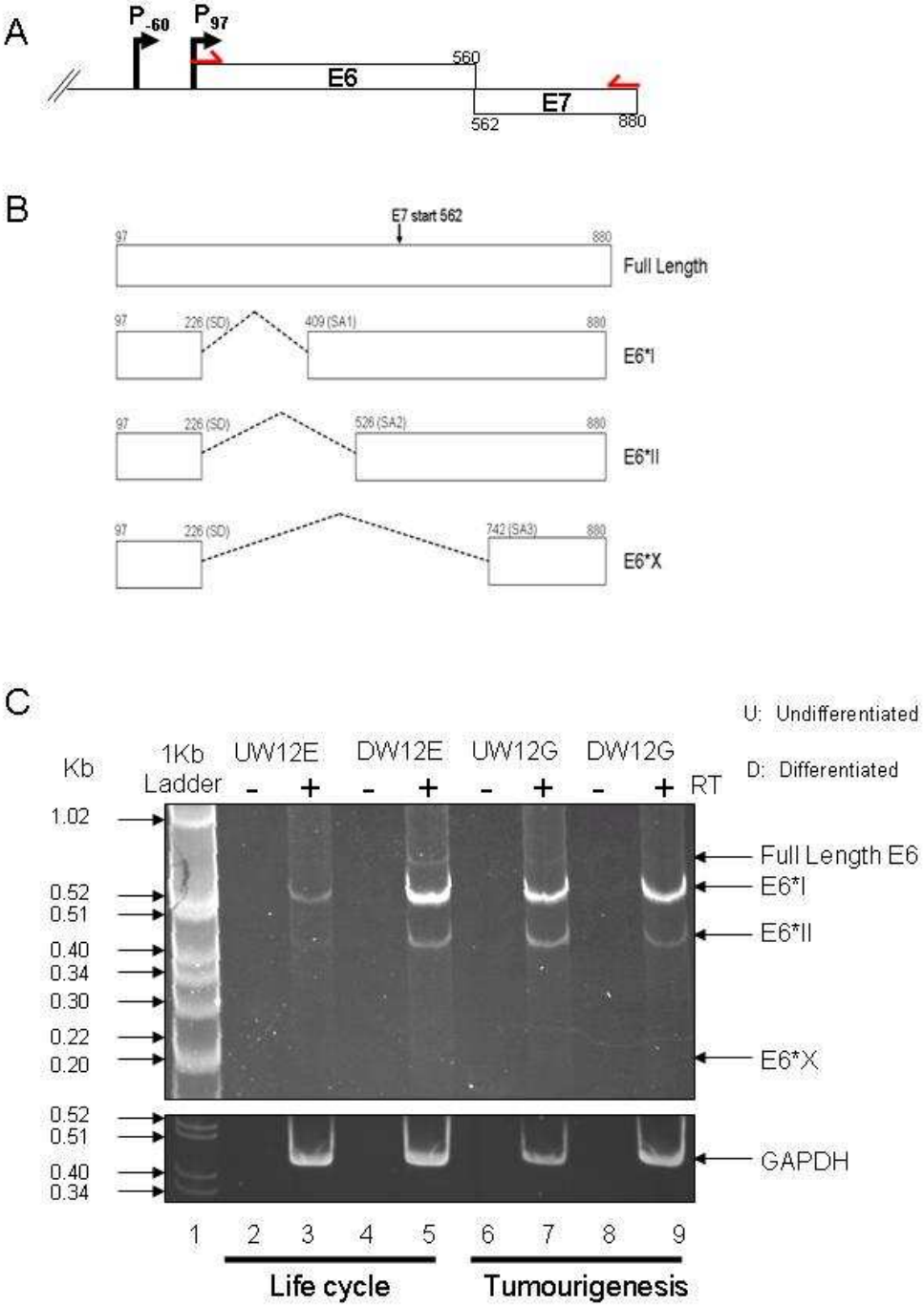
3.1.1 Alternative splicing of E6 during epithelial differentiation

HPV16 E6/E7 gene transcripts can be alternatively spliced and the putative splice isoforms are shown in Figure 3.2 B. With the production of multiple RNA isoforms there are two hypotheses I addressed. First, E6 isoform levels may change during the virus replication cycle or during tumour progression if a function attributed to a particular isoform is only required under certain conditions. Second, any proteins produced by these alternatively spliced RNAs may have differing functions. In the first experiment shown in Figure 3.2, I addressed the first hypothesis. In order to determine whether alternative splicing of HPV16 E6/E7 is altered during the virus life cycle, RNA was extracted from W12E cells which were originally isolated from a low grade cervical lesion and contain HPV16 episomal genomes. (Stanley et al., 1989, Jeon et al., 1995). These cells provide a model system for the study of the viral life cycle as they are capable of epithelial differentiation: they produce virus particles. As a control for productive viral replication I also isolated RNA from W12G cells. W12G cells (clone 20861) were cloned from the original W12 cell line at the same time as W12E cells (clone 20863) and contain multiply integrated HPV16 genomes with no episomal genomes. These cells have more limited differentiation capacity and are incapable of virus production.

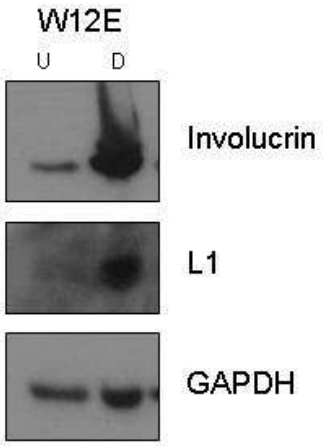
Previous work in our laboratory has identified the optimal days for harvesting cells in different populations and obtaining practical concentrations of starting material. McPhillips et al (2004) showed that the differentiation level of the

W12E and G cell populations at day 5 is low as indicated by low levels of the differentiation marker involucrin at day 5 while at day 10 high levels of involucrin were detected (McPhillips et al., 2004). Similar results were found using confocal microscopy. At day 5 70-85% of the cell population is undifferentiated while at day 10, 70-80% of the cell population is differentiated (Milligan et al., 2007). Markers of cellular differentiation such as involucrin, keratin 10 and filaggrin can be readily detected (Cheunim et al., 2008). An example of the involucrin protein levels extracted from the same W12E cells at day 5 and day 11 are shown in Figure 3.2 D. This experiment was carried out by Dr T. Klymenko using the same cell stocks I have used in my experiments. There is a clear increase in involucrin protein levels in the day 11 extract compared to the day 5 extract. Involucrin is detected at a low level in the undifferentiated cell population because these are a heterogeneous cell population where a small percentage of the cells will have begun to differentiate. W12G cells also showed increased involucrin expression, although not to the same level as in W12E cells (data not shown). For the viral life cycle, the late viral protein E1^E4 was not detected after W12 cells were grown for five days in cell culture but was detected after ten days (McPhillips et al., 2004). Milligan et al (2007) showed that after 10 days of culturing, W12E cells could be found to be expressing L1 protein in a subset of cells (Milligan et al., 2007). This result indicates that after ten days of culture the W12 cells are differentiated at least to the extent of allowing viral capsid protein formation. An example of L1 protein levels at day 5 and day 11 in the same W12E cells as used in this experiment is shown in Figure 3.2 D. Here there is a clear increase in L1 protein levels in the day 11 extract when compared to the day 5 extract. GAPDH levels were used as an internal control standard.

The W12E and G RNAs were reverse transcribed followed by PCR amplification using primers that amplify the whole E6/E7 open reading frame (red arrows Figure 3.2 A, Table 2). Undifferentiated W12E and W12G cells were grown in an identical manner on J2 3T3 fibroblast feeder layers at low density for five days before the feeder layers were removed and the cells harvested for RNA preparation. Differentiated W12E and W12G cells were grown on the same feeder layer cells to high density for ten days before the feeder layer cells were removed and the W12 cells harvested for RNA extraction. All cell populations were grown in DMEM medium containing 1.2 mM Ca^{++}: differentiation was controlled by cell density alone as described previously (McPhillips et al., 2004). Polyadenylated RNA was extracted from total RNA. The polyadenylated RNA was DNase-1 treated before being reverse transcribed and amplified by semi-quantitative PCR using GAPDH as an internal control standard. The PCR products were resolved on 6% acrylamide gels and stained with ethidium bromide and viewed under UV. Figure 3.2 C shows the result of acrylamide gel electorphoretic separation of E6 isoforms from undifferentiated and differentiated W12E and G cells. Tracks 2, 4, 6, and 8 show the reaction carried out in the absence of reverse transcriptase. These control tracks are designed to demonstrate the absence of DNA in the input RNA samples in the cDNA reaction. No bands were observed in any of these tracks. Track 3 shows that only low levels of E6


A
P-60
P97
E6
560
E7
562
880
B
E7 start 562
97
880
Full Length
97
226 (SD)
409 (SA1)
880
E6*I
97
226 (SD)
526 (SA2)
880
E6*II
97
226 (SD)
742 (SA3)
880
E6*X
C
U: Undifferentiated
D: Differentiated
Kb
1Kb Ladder
UW12E
DW12E
UW12G
DW12G
- + - + - + - + RT
1.02
0.52
0.51
0.40
0.34
0.30
0.22
0.20
0.52
0.51
0.40
0.34
Full Length E6
E6*I
E6*II
E6*X
GAPDH
1 2 3 4 5 6 7 8 9
Life cycle
Tumourigenesis

D


W12E
U
D
Involucrin
L1
GAPDH

E

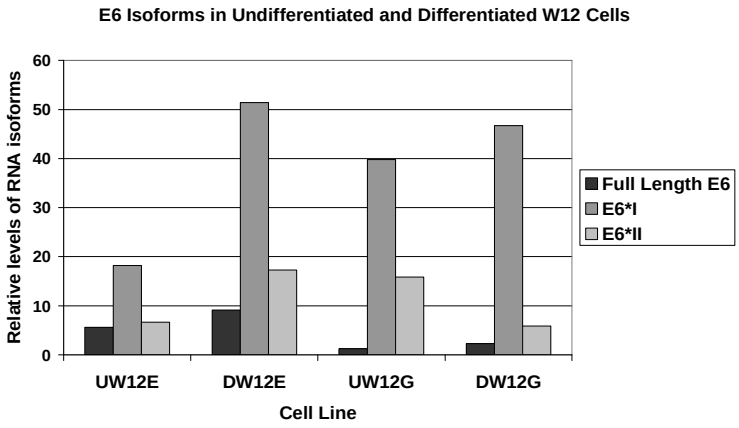

E6 Isoforms in Undifferentiated and Differentiated W12 Cells
Relative levels of RNA isoforms
60
50
40
30
20
10
0
UW12E
DW12E
UW12G
DW12G
Cell Line
Full Length E6
E6*I
E6*II

Figure 3.2 E6 isoform expression is not altered upon epithelial differentiation. **A)** Schematic diagram depicting the E6/E7 open reading frame. The position of the early promoters, P-60 and P97 is shown by the black arrows and the position of the E6/E7 primers used in this study is shown by red chevrons. **B)** Schematic diagram of the E6/E7 isoforms with the splice donor (SD) and acceptor (SA) sites labelled and the nucleotide numbers given. **C)** Upper panel Ethidium-bromide stained acrylamide gel electrophoresis of E6/E7 isoforms RT-PCR amplified from polyadenylated W12 RNA during epithelial differentiation. Isoforms are indicated to the right of the gel. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "UW12E", undifferentiated W12E RNA; "DW12E", differentiated W12E RNA; "UW12G", undifferentiated W12G RNA; "DW12E", differentiated W12E RNA. Lower Panel Ethidium-bromide stained acrylamide gel electrophoresis of GAPDH RNA RT-PCR amplified from polyadenylated W12 RNA during epithelial differentiation. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "UW12E", undifferentiated W12E RNA; "DW12E", differentiated W12E RNA; "UW12G", undifferentiated W12G RNA; "DW12E", differentiated W12E RNA. **D)** Western Blot analysis of protein extracted from undifferentiated (U) and differentiated (D) W12E cells. The protein being detected is indicated to the right of the gel. Involucin is a marker for differentiation, L1 is a late viral protein and GAPDH is used as a loading control. **E)** Quantification of the E6 isoforms shown in Figure 3.1 (C). The gel was scanned and band density estimated using Photoshop. The levels of E6 isoforms are calculated relative to GAPDH mRNA levels. The results shown are from one experiment. Two experiments were carried out with very similar results.

isoforms could be detected in undifferentiated W12E cells where there is detectable E6*I and E6*II mRNAs. There is no detectable E6 full length or E6*X RNA expression suggesting that these isoforms have very limited expression in those cells. Track 5 shows the E6 isoforms detected in differentiated W12E cells. Similar to undifferentiated W12E cells (track 3) E6*I is the predominant isoform however the total levels of E6/E7 RNAs appeared to increase significantly upon epithelial differentiation and E6 full length and E6*II were now detectable. E6*X was not detected. Both experiments were carried out with similar starting concentrations of RNA therefore the increased levels of E6/E7 RNAs are differentiation-dependent. To test whether there might be a requirement for higher levels of E6 and E7 as the viral life cycle progresses I carried out a similar experiment in W12G cells which have only integrated HPV16 genomes and where there is no virus life cycle completion. Track 7 shows the E6 isoforms detected in undifferentiated W12G cells. Similar to W12E cells, full length E6, E6*I and E6*II mRNAs were all detectable in those cells. Again E6*X was not detected in undifferentiated W12G cells. Track 9 shows the E6 isoforms detected in differentiated W12G cells. Again the three largest isoforms were detected while the smallest isoform, E6*X was not. In contrast to the findings from W12E cells however, the total levels of E6 RNAs were not increased upon differentiation of W12G cells yet although W12G cells display a reduced differentiation capacity, they were higher than the levels in undifferentiated W12E cells. As W12G cells contain integrated HPV16 genomes the expression levels of E6 and E7 RNAs and proteins may be higher than in W12E cells where E2 is present to control the expression of E6 and E7. Indeed, Northern blot analysis of E6/E7 mRNA levels comparing W12E with W12G cells indicated around five-fold more E6/E7 mRNA in W12G cells that W12E cells (SV Graham, unpublished observations). The experiment in Figure 3.2 C was carried out twice with very

similar results. Figure 3.2 E shows a quantification of one of the experiments. The data were quantified by measuring the pixel density of each of the individual isoforms relative to GAPDH levels. Combination of the data from the two independent experiments was not possible as the raw figures were vastly different and masked any difference that was visible from the raw data. Although the pixel density numbers were different the trend in the data was not and both experiments showed similar results. Quantitative PCR experiments were also attempted to quantify properly the levels of each of the E6 isoforms during differentiation of the epithelial cells. However Taqman probe and primer design proved impossible for the individual isoforms. The only way to specifically amplify individual isoforms was to design either one of the primers or the probe to cover the splice junction as this is the only part of the RNAs that is unique to the individual isoforms. Unfortunately the area surrounding the splice junctions was not optimal for primer design. Initial attempts to quantify the isoforms did not work as the probe and primer sets were not compatible with the relative quantification method; the primer efficiencies were not equal or even similar. Quantitative PCR is an excellent method for quantification if the procedure is carried out correctly, however the E6 isoforms proved unsuitable for this type of quantification. The probe and primer efficiencies are shown in Table 6.

Probe and Primer Set	Efficiency
E6 Full Length	92%
E6*I	94%
E6*II	86%
E6*X	85%
B-Actin	99%

Table 6: Efficiencies of E6 isoform probe and primer sets. The efficiencies of the E6 probe and primer sets were calculated using a standard curve of W12GPXY cDNA. For probe and primer sets to be compatible for relative quantification they should have efficiencies of +/- 10% with the endogenous reference gene (β-Actin).

Contrary to what was suggested by data in the recent report by Rosenberger et al (2010), I did not detect any alteration in E6 isoform RNA expression during differentiation of the W12E or W12G epithelial cells (Rosenberger et al., 2010). Rosenberger et al (2010) predicted that due to the relatively higher levels of EGF in the basal epithelial layer there should be higher levels of full length E6 RNAs than in the upper layers (Rosenberger et al., 2010). However, in my experiments differentiation of the epithelial cells increased the levels of all E6 transcripts which could suggest that differentiation of the epithelial cells promotes transcription from the P_{97} promoter. This may or may not be in response to

changing EGF levels. It should also be noted that EGF is added to the W12 cell media every two days therefore EGF levels may remain high throughout the epithelium. Rosenberger et al (2010) did not use W12 cells but used immortalised keratinocytes stably transfected with either the full HPV16 genome or E6/E7 alone. This means they were examining E6/E7 expression under conditions of over expression and when carrying out experiments with "low" EGF this was achieved by not adding EGF to the media (Rosenberger et al., 2010). The increase in levels of E6 transcripts that I observed upon epithelial differentiation may suggest an increased requirement for E6 and E7 activities in differentiated cells, however the increase may also be because there is an increased requirement for other viral proteins and because the viral RNAs are polycistronic E6 and E7 are also transcribed at higher levels. The cooperative and antagonistic effects of E6 and E6*I may allow this to occur without resulting in catastrophe for the cell as any unwanted E6 actions could be prevented by interaction with E6*I. It is not surprising that the increase in E6/E7 mRNA levels was only identified in W12E cells and not in W12G cells. HPV16 DNA is episomal in W12E cells. This means that the virus is able to go through a productive replication cycle, where the demand for expression of individual viral proteins is altered depending on the differentiation status of the infected cell. Whereas in W12G cells the viral DNA is integrated and there is no virus replication taking place and therefore there is no requirement for altered viral protein levels. As there is no productive viral replication in W12G cells and the viral genome is integrated into the host cell genome, expression of E6/E7 RNAs will be controlled by cellular transcription factors.

In conclusion, this experiment indicated that E6 splicing is not changed upon epithelial differentiation. However the requirement for E6 appears altered during epithelial differentiation. The levels of all of the detectable E6 isoforms are increased upon differentiation suggesting the virus requires increased levels of E6 in later stages of the viral life cycle. The increase in E6 levels is not due to any non-specific epithelial differentiation signals as a similar increase in E6 levels is not seen upon differentiation of W12G cells where there is no productive viral replication taking place. It is not clear however, from this experiment, whether any proteins are produced by the RNA isoforms and whether they might have different functions.

3.1.2 Alternative splicing of E6 during cervical tumour progression

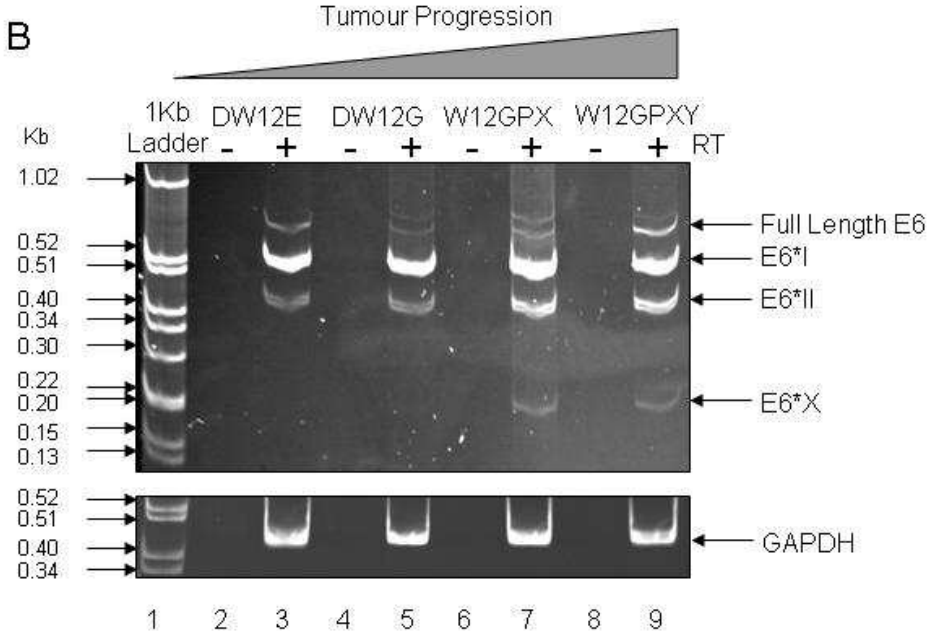
In order to determine whether alternative splicing of HPV16 E6/E7 is altered during HPV-induced transformation of infected epithelial cells RNA was extracted from the four W12 cell lines that model HPV16-induced cellular transformation (Stanley et al., 1989, Jeon et al., 1995, Aasen et al., 2003). The W12 cervical cancer progression model is depicted in Figure 3.2 A. The RNA was reverse transcribed followed by PCR amplification using the same primers used in section 3.1.1 and illustrated in Figure 3.2 A. For this experiment differentiated populations of W12E and W12G cells were chosen purely to maximise the RNA concentrations obtained from the extractions and as shown in Figure 3.2 C differentiation had no effect on the pattern of alternative splicing of the E6/E7 RNAs. W12GPX and W12GPXY cells did not require feeder cells for

growth so these were simply harvested directly for RNA extraction. Polyadenylated RNA was selected from total RNA. All polyadenylated RNAs were DNase-1 treated before being reverse transcribed and amplified in a semi-quantitative PCR reaction using GAPDH as an internal control for input of equal concentrations of RNA to the reactions. The PCR products were resolved on a 6% acrylamide gel and stained with ethidium bromide. Again acrylamide gels were used as they provide superior separation of DNA with molecular weight of less than 1Kb. The amplified products are shown in Figure 3.3 B. Tracks 2, 4, 6 and 8 show the reactions carried out in the absence of reverse transcriptase which confirms the absence of DNA from the RNA starting material for the cDNA reaction. Similar to what was observed in Figure 3.2 C there was no apparent difference in E6 isoform production between W12E (track 3) and W12G (track 5) cells. However, several differences between the transformed cell lines, W12GPX (track 7) and W12GPXY (track 9), and the untransformed W12E and W12G cells were noted. W12GPX and W12GPXY cells expressed all four E6/E7 isoforms including E6*X which was not detectable in W12E (track 3) and W12G cells (track 5). The most highly expressed isoform in all samples was E6*I, in agreement with the original findings Smotkin and Wettstein (Smotkin & Wettstein, 1986, Smotkin et al., 1989). The appearance of E6*X only in the transformed cells could indicate that the smallest spliced isoform, is a tumour specific isoform and may be involved in the transformation of the epithelial

A

W12E	W12G	W12GPX	W12GPXY
Episomal genome	Integrated genome	Integrated genome	Integrated genome
Immortal	Immortal	Transformed	Transformed
Can Differentiate	Can Differentiate	No Differentiation	No Differentiation
No invasion	No invasion	Invasion	Extensive Invasion
Not tested *in vivo*	Forms cysts in mice	Not tested *in vivo*	Forms SCC tumours in mice

B



C

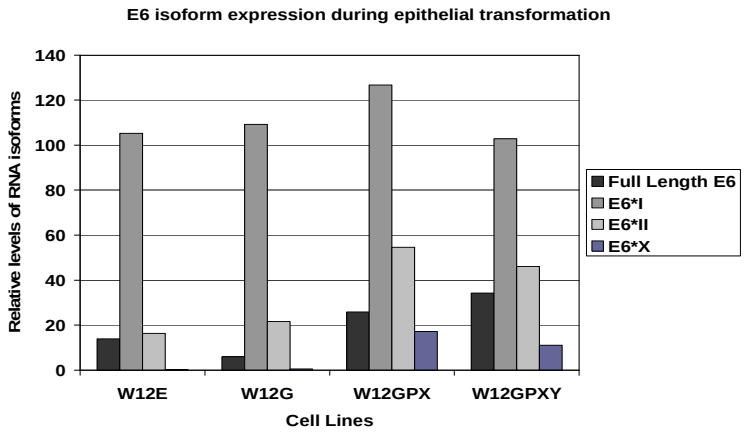


Figure 3.3 E6 isoform expression is altered upon epithelial transformation. **A)** Table listing the properties of the W12 model cell lines. **B)** Upper panel Ethidium-bromide stained acrylamide gel electrophoresis of E6/E7 isoforms RT-PCR amplified from polyadenylated W12 RNAs during epithelial transformation. Isoforms are indicated to the right of the gel. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "DW12E", differentiated W12E RNA; "DW12G", differentiated W12G RNA; "W12GPX", W12GPX RNA; "W12GPXY", W12GPXY RNA. Lower panel Ethidium-bromide stained acrylamide gel electrophoresis of GAPDH RNA RT-PCR amplified from polyadenylated W12 RNAs during epithelial transformation. Isoforms are indicated to the right of the gel. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "DW12E", differentiated W12E RNA; "DW12G", differentiated W12G RNA; "W12GPX", W12GPX RNA; "W12GPXY", W12GPXY RNA. **C)** Quantification of the E6 isoforms shown in Figure 3.2 (B). The gel was scanned and band density estimated using Photoshop. The levels of E6 isoforms are relative to GAPDH mRNA levels. The results shown are from one experiment. Three experiments were carried out all with very similar results.

cells. Similarly the E6*II isoform appeared to be increased upon epithelial transformation suggesting that there is perhaps a shift towards expression of the smaller isoforms. The experiment in Figure 3.3 B was carried out three times with similar results. Figure 3.3 D shows a quantification of one such experiment. The data were quantified by measuring the pixel density of each of the individual isoforms relative to GAPDH levels. Combination of the data from the three independent experiments was again not possible due to the same reasons as described in section 3.1.1. Although the pixel density numbers were different the trends in the data were very similar. Quantitative PCR experiments were also attempted to quantify properly the levels of each of the E6 isoforms during differentiation of the epithelial cells. However the same difficulties were encountered as described in section 3.1.1. This gel also appears to show doublet bands for most of the isoforms however this is probably just a feature of this particular gel as doublet bands are not seen in every gel. Similarly in this gel there is a band higher than full length E6 which could be a second full length E6 isoform due to the promoter reported to be present upstream from the major early promoter (Milligan et al., 2007). This band would have to be confirmed by cloning and sequenceing, which was attempted but due to the low levels of expression, this was never confirmed.

I conclude that E6 RNA isoform expression is altered upon transformation of the epithelial cells. The smaller isoforms appear to be upregulated in transformed epithelial cells which suggests that proteins produced by the RNA isoforms may have differing functions. The upregulation of the smaller isoforms during transformation of the virus-infected epithelial cells suggests these isoforms may

have growth promoting or tumour promoting functions that are normally tightly regulated during a productive viral infection.

Next, it was important to identify whether the increased expression of the smallest isoform E6*X was unique to the W12 transformed cell lines or was a general feature of HPV16-positive cervical cancer cell lines. It was possible that as the GPX and GPXY cells were derived in a laboratory setting that this splicing event may not be present in other, more directly derived cervical tumour cell lines. Therefore, RNA was harvested from two other HPV16 positive cervical cancer cell lines that were derived completely independently from the W12 cell lines and originated from high grade cervical tumours, CaSki and SiHa cells. The RNA was DNase-1 treated before being reverse transcribed and amplified by semi-quantitative PCR using the E6 primers described in section 3.1.1. GAPDH was used as an internal control standard. The PCR products were resolved on 6% acrylamide gels and stained with ethidium bromide and viewed under UV. Figure 3.4 shows the result of acrylamide gel electorphoretic separation of E6 isoforms from the cervical tumour-derived cells and from W12GPXY cells. Tracks 2, 4 and 6 show the reactions carried out in the absence of reverse transcriptase confirming the absence of DNA from the RNA used in the cDNA reactions. Track 3 shows the E6/E7 mRNA isoforms expressed in SiHa cells. Full length E6, E6*I and E6*II can all be detected. Track 5 shows the mRNA isoforms expressed in W12GPXY cells which, as shown in Figure 3.3 C, expressed full length E6, E6*I, E6*II and E6*X. Track 7 shows the mRNA isoforms expressed in CaSki cells, which like GPXY cells express all four E6 isoforms, however E6*X is expressed at very low levels and can only be seen when the gel is over-exposed. The experiment

was carried out three times with very similar results. This experiment indicated that the increased levels of E6*X compared to other E6 isoforms is not unique to W12 cells and that the isoform is also present at increased levels in CaSki cells. Very similar amounts of RNA starting material were used in each sample. The difference in levels of E6/E7 RNAs between the different tracks (tracks 3, 5 and 7) is likely due to the number of HPV16 genome insertions in each cell line. CaSki cells have many more genome insertions (more than 500) than SiHa which reported has less than 10 genome insertions (Yee et al., 1985) and the lack of detectable E6*X in SiHa cells is likely due to the limits of detection and not due to a lack of isoform expression.

For this experiment I conclude that the alteration in E6 splicing observed in W12GPX and W12GPXY cells is not unique to these cell lines. E6*X is also detectable in epithelial tumour-derived CaSki cells. It is not possible to say whether E6*II is also upregulated in these cells, as there is no non-transformed cells for comparison. These findings suggest a probable tumorigenic property for E6*X and perhaps E6*II.

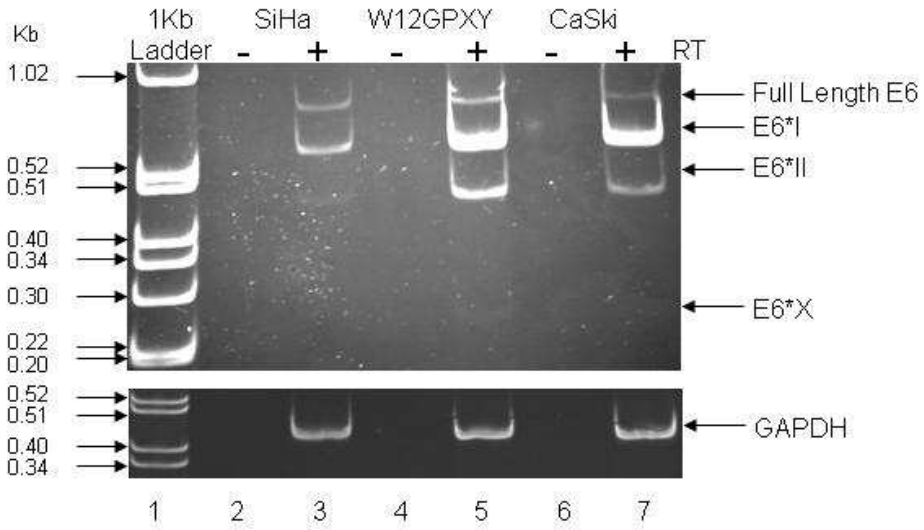


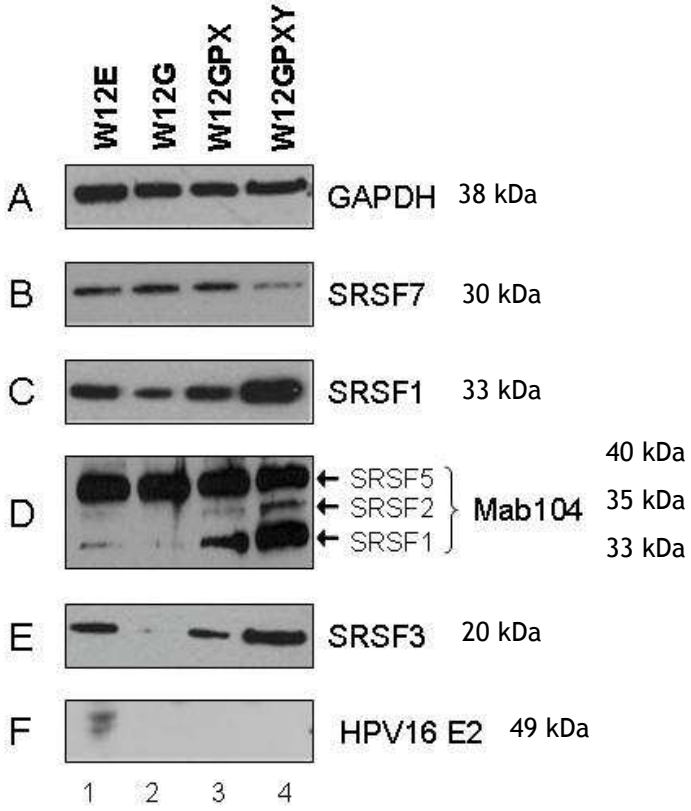
Figure 3.4 E6 isoforms are also expressed in other HPV16 transformed epithelial cells lines Upper Panel Ethidium-bromide stained acrylamide gel electrophoresis of E6/E7 isoforms RT-PCR amplified from SiHa, W12GPXY and CaSki RNA. Isoforms are indicated to the right of the gel. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase. Lower Panel Ethidium-bromide stained acrylamide gel electrophoresis of GAPDH RNA RT-PCR amplified from SiHa, W12GPXY and CaSki RNA. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase.

3.1.3 SR proteins levels are altered upon transformation

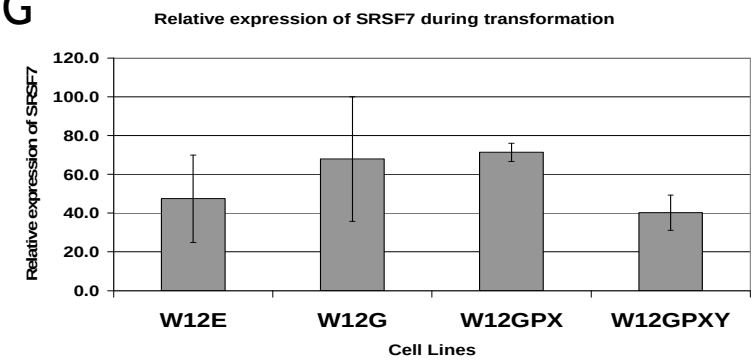
Cellular transformation often results in altered gene expression. Transformation often leads to genomic instability and with this, control over gene expression and DNA replication is lost. For example during transformation control over expression of oncogenes including c-Myc can be lost resulting in increased proliferation and resistance to apoptosis. Conversely tumour suppressor functions are often lost during transformation. For example BRCA1 tumour suppressor is often mutated in breast cancers which interferes with the DNA damage repair pathway and allows a build up of genetic mutations leading to transformation. Lack of control during DNA replication can result in alteration in the chromosome copy numbers leading to an increase in expression of genes encoded on the affected chromosomes. In particular, recently several groups have demonstrated alterations in SR protein expression in various tumours such as breast and ovarian cancers (Stickeler et al., 1999, Fischer et al., 2004). SR proteins are the cellular splicing factors that are involved in determining which RNAs are to be spliced and alternative splicing is regulated by SR proteins (reviewed by (Long & Caceres, 2009). The changes in SR protein expression noted in tumour cells likely have functional significance because Karni et al (2007) identified that an oncogenic isoform of S6K1 is often overexpressed in breast and lung cancer cells after overexpression of SRSF1 (Karni et al., 2007). This S6K1 isoform was then shown to be sufficient for the transformed phenotype of the cells. Overexpression of S6K1 isoform 2 (the oncogenic isoform) was able to promote anchorage independent growth in NIH 3T3 cells (Karni et al., 2007). Based on these observations I hypothesised that SR proteins may be overexpressed in cervical cancers resulting in changes in HPV

oncoprotein isoform expression during cervical tumour progression, especially if smaller isoforms such as E6*I had antagonistic properties to E6 full length.

The preceding experiments demonstrated that E6/E7 expression is altered upon cellular transformation due to a change in the splicing pattern of HPV16 E6/E7 mRNAs. Therefore, the next step was to investigate any changes in the levels of the SR splicing regulator proteins in the transformed cells. Total cellular protein was harvested from differentiated W12E and W12G cells and W12GPX and W12GPXY cells. I chose to examine differentiated W12E and G cells because we had demonstrated previously increase expression of SR proteins upon HPV-16 positive cervical epithelial differentiation (Mole et al., 2009b). The proteins were electrophoresed on a gradient 4-12% acrylamide gel in SDS denaturing conditions and blotted onto a nitrocellulose membrane. The membrane was then western blotted with SR protein-specific antibodies. Figure 3.5 B-E shows western blot analysis of relative levels of each of the indicated SR proteins. The protein being detected is indicated on the right of each membrane. GAPDH was used as an internal control standard for protein loading (Figure 3.5 A). These experiments were carried out three times and the protein levels quantified relative to GAPDH (Figure 3.5 G-J). In the western blot panels, track 1 (W12E) shows that SRSF1 and SRSF3 are all expressed to higher levels than in track 2 (W12G) whereas other SR proteins, for example, SRSF5 and SRSF7 are expressed at very similar levels. Western blot panel D was probed with Mab104. This antibody detects all classical SR proteins. Only the portion of the blot containing SRSF1, 2 and 5 is shown. Close examination of the bands corresponding to SRSF2 also indicates an increase in expression comparing W12E to W12G cells.

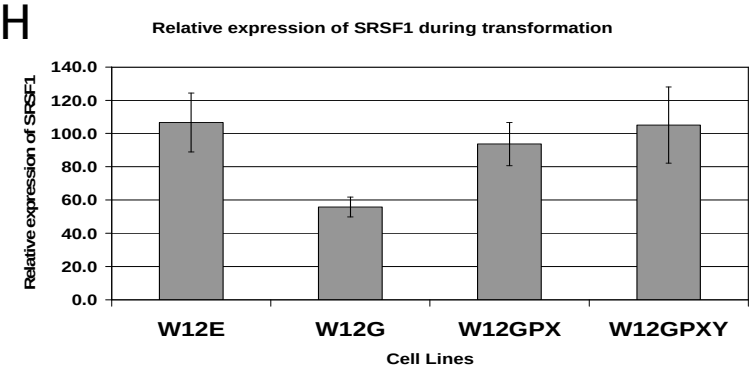

W12E
W12G
W12GPX
W12GPXY
A
GAPDH
38 kDa
B
SRSF7
30 kDa
C
SRSF1
33 kDa
D
SRSF5
SRSF2
SRSF1
Mab104
40 kDa
35 kDa
33 kDa
E
SRSF3
20 kDa
F
HPV16 E2
49 kDa
1 2 3 4

G



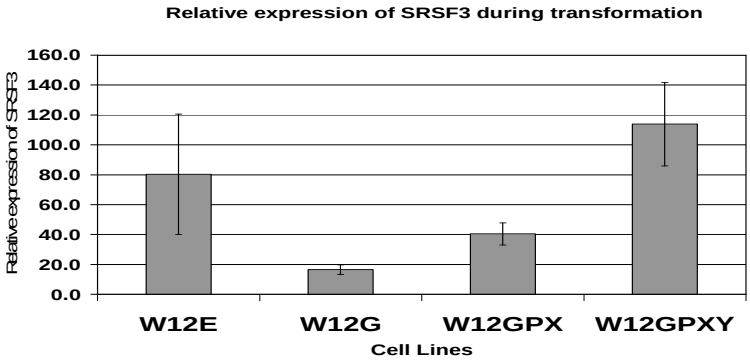
SRSF7

H



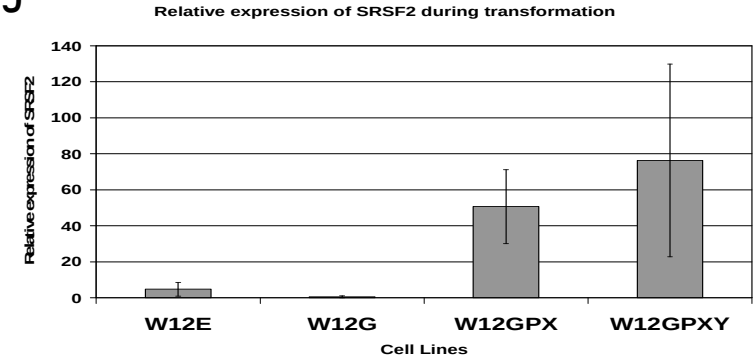
SRSF1

I



SRSF3

J



SRSF2

Figure 3.5 Specific SR proteins are upregulated upon cervical epithelial transformation. Western Blot analysis of SR protein levels in the various W12 model cell lines. The cell lines are indicated to the top of the blots. **A)** Blot probed with Mab6CS against GAPDH and used as an internal control. **B)** Blot probed with Mab98 showing SRSF7 levels in the model cell lines. **C)** Blot probed with Mab96 showing SRSF1 levels in each cell line. **D)** Blot probed with Mab104 showing SRSF1, SRSF2 and SRSF5 levels in the model cell lines. Only the portion of the gel showing SRSF1, 2 and 5 is shown. **E)** Blot probed Mab7B4 showing levels of SRSF3 protein in the W12 model cell lines. **F)** Blot probed with TVG261 showing HPV16 E2 levels in the W12 model cell lines. Quantification of three independent experiments showing the mean and standard deviation for three experiments is shown in the graphs. In every case the levels of SR proteins are calculated relative to GAPDH levels. **G)** SRSF7 levels **H)** SRSF1 levels **I)** SRSF2 levels and **J)** SRSF3 levels

Unfortunately, the antibodies specific for this SR protein do not work efficiently in western blotting so I was unable to obtain better evidence for changes in SRSF2 expression between these two cell lines. This suggests that SRSF 1-3 are upregulated either by differentiation or by full viral protein expression in W12E cells where the viral DNA is episomal and is able to support the virus life cycle. For SRSF1 and SRSF3 our laboratory has shown upregulation to be a result of HPV16 E2 transcriptional control of the genes encoding these proteins (Mole et al., 2009a, Mole et al., 2009b). When levels of the SR proteins in tracks 2, 3 and 4 are compared an increase in levels of the same three proteins can be observed, particularly between W12G (track 2) and W12GPX cells (track 3). The viral genome is integrated into the host DNA in these cells and the E2 protein is not expressed (Figure 3.5 F). Therefore E2 cannot be responsible for the observed upregulation in the tumour progression cell models. The proteins are being upregulated by some other mechanism. The overexpression of SRSF1, 2 and 3 is specific to these proteins as other SR proteins, for example SRSF5 and SRSF7, remained relatively constant in all cell lines regardless of whether the cells were transformed or had a productive viral infection (Figure 3.5 G-J).

Further experiments to determine if the SRSF protein overexpression also occurs in patient tissue were undertaken. Cervical biopsy samples stored in paraffin were stained for the presence of SR proteins SRSF1-3. Diagnosis of the severity of the cervical lesion was performed by Dr David Millan a gynaecological pathologist at Glasgow Royal Infirmary. Tissues were stained by the Institute of Comparative Medicine's Histopathology service at the University of Glasgow. Figure 3.6 Shows representative images from 10 CIN 1 and 10 CIN3 lesions. The protein being detected is indicated to the right of the images. In all cases expression of the SR protein in low grade (CIN1) lesions is found in the lower and

middle layers of the epithelium with little expression in the terminally differentiated cells at the top of the epithelium. However in CIN 3 lesions SR proteins were detected throughout the epithelium. This is probably because the epithelium is now filled with basal-like cells that are no longer differentiating. This confirms the western blotting experiments in Figure 3.5. The SR protein overexpression seen in W12GPX and W12GPXY cells is due to the presence of basal like cells with no differentiation taking place.

The results indicate that some (SRSF1, SRSF2 and SRSF3) but not all, SR proteins are specifically upregulated upon transformation of W12 epithelial cells and in patient tissue. This upregulation of SR proteins could explain the changes in splicing of the E6/E7 RNAs upon W12 cell transformation. It is also clear that the same SR proteins, SRSF1, SRSF2 and SRSF3, are upregulated in W12E cells when compared to W12G cells. This is in agreement with the recent publication by Mole et al (2009) which reported that HPV16 E2 could bind and transcriptionally activate the SRSF1 promoter resulting in an increase of SRSF1 protein. SRSF1 has been shown to be required to process viral E4^L1 RNAs at least from subgenomic constructs in HeLa cells (Somberg & Schwartz, 2010). More recent studies from our laboratory have demonstrated E2 regulation of SRSF2 and 3 (Mole et al., 2009b). The reduced levels of SRSFs 1-3 in W12G cells are likely due to the absence of E2 due to the viral genome being integrated into the cellular DNA and disrupting the E2 open reading frame.

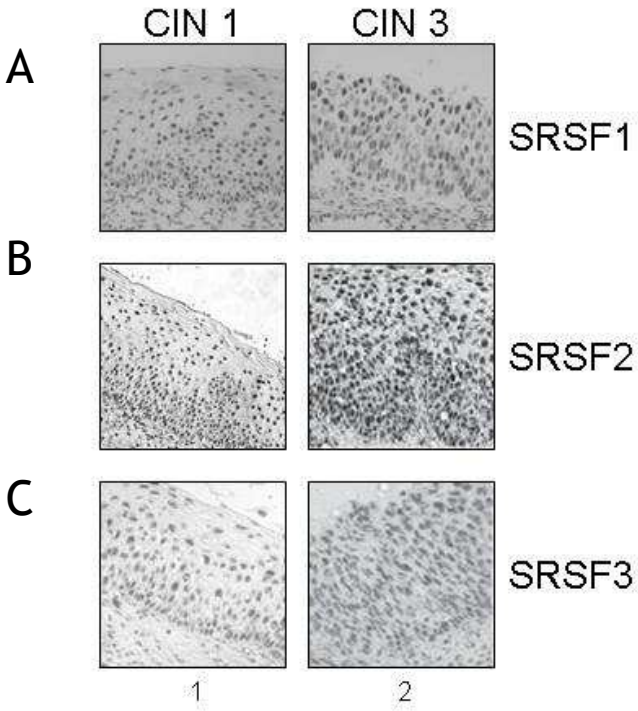


Figure 3.6 Specific SR proteins are upregulated in cervical lesions. Immunohistochemical analysis of SR protein levels in CIN1 and CIN3 cervical biopsies. The lesion severity is indicated at the top of the pictures. **A)** Cervical lesions probed with Mab96 against SRSF1. **B)** Cervical lesions probed with Mab SC-35 showing SRSF2 levels in the cervical biopsies. **C)** Cervical lesions probed with Mab7B4 showing SRSF3 levels in the cervical biopsies.

To conclude, epithelial transformation by HPV causes an upregulation of specific members of the SR protein splicing regulatory family. This upregulation probably aids in the transformation process by resulting in a number of alterations in splicing of cellular RNAs. Changes in alternative splicing will lead to production of a tumour specific-proteome.

3.1.4 SR protein overexpression may not be due to chromosomal duplication

In the W12 transformed cells the increase in SR protein levels identified in section 3.1.3 cannot be due to the presence of HPV16 E2 because it is not expressed in these cells. Therefore there must be a different mechanism involved in the upregulation of SR proteins during cellular transformation. One possible reason for this upregulation could be that the SR genes become amplified during cellular transformation. This has been shown to be the case in some breast cancers as Karni et al (2007) showed that the 17q23 chromosome containing SRSF1 becomes duplicated in some breast cancer cells (Karni et al., 2007). In order to test whether this is the case in our W12 model cell lines DNA was extracted from differentiated W12E and W12G cells and W12GPX and W12GPXY cells and qPCR carried out in order to determine the gene copy number of the SR gene in each case. DNA was extracted and diluted to equal concentrations for each cell line. qPCR was carried out using Stratagene Mastermix and Applied Biosystems commercially available predesigned Taqman probe and primer sets. All qPCR experiments were carried out using the Applied Biosystems 7500 machine and the data analysed using the 7500 system software.

Standard curves of W12E DNA were generated using the Applied Biosystems user manual instructions. W12E DNA was used to create the standard curve as the viral DNA is episomal and is unlikely to have caused significant alteration to the cellular chromosomes. Similar to some breast cancers a significant difference in SRSF1 gene copy numbers was identified in the assay comparing SRSF1 gene copy number in W12E with W12GPXY cells. W12E cells gave a CT value of 29 ± 1 while W12GPXY gave 25 ± 1. This difference was significant because a student's T-test gave a p value of 0.029 (Figure 3.7, asterisk). In CT values a difference of 1 is actually a two-fold difference as one CT value is a doubling of the target. Therefore a difference of 4 is actually a 16-fold difference in target DNA. This indicates that there has been an increase in the gene copy number of the SRSF1 gene between derivation of the W12GPXY cell line from W12GPX epithelial cells. Figure 3.5 C showed that there was significantly more SRSF1 protein detected in W12GPX cells compared to W12G cells yet I discovered no significant difference between the SRSF gene copy number in these cell lines. This indicates that increased levels of SRSF1 protein in W12GPX cells must be due to another mechanism. Interestingly, I was not able to detect any significant difference between any of the cell lines when qPCR was used to amplify the SRSF2 or SRSF3 genes (Figure 3.7) indicating that there is transcriptional upregulation or an increase in protein stability of these proteins in the transformed cells.

I conclude that transformation of the W12 epithelial cells leads to a specific overexpression of SRSF1, 2 and 3. The overexpression of SRSF1 is due to an increase in gene copy number at least in part, similar to breast cancer cells. The SRSF2 and SRSF3 overexpression cannot be explained by gene amplification therefore the upregulation is most likely due to transcriptional upregulation of the SR RNAs and proteins or an increase in the stability of the proteins.

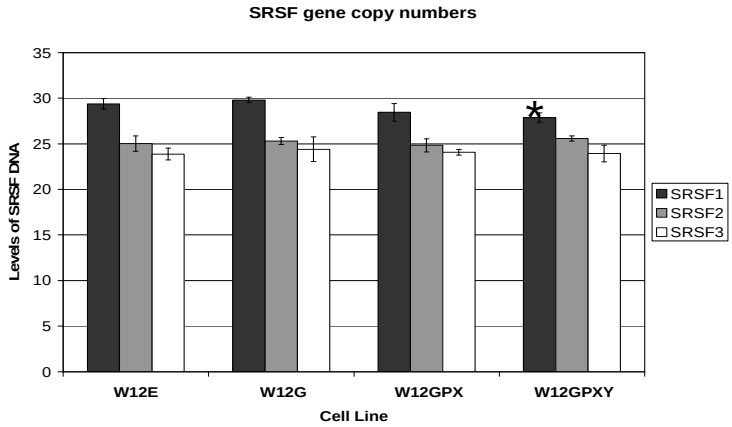


Figure 3.7 SRSF1 DNA is amplified during cervical epithelial cell transformation but SRSF2 and SRSF3 DNA is not. qPCR analysis of DNA copy numbers in the W12 model cell lines. The experiment was carried out three times and the graph shows the mean and standard deviation of three independent experiments. SRSF1 DNA is significantly amplified in W12GPXY cells indicated by "*". SRSF2 and SRSF3 DNA are not significantly different.

3.1.5 HPV11 E2 also up-regulates SRSF3 expression

As certain SR proteins are overexpressed in differentiated W12E cells and this correlate with HPV16 E2 expression (Figure 3.5) and it has previously been shown that HPV16 E2 activates the promoter of the genes encoding SRSF1 and 3 (Mole et al., 2009a, Mole et al., 2009b), I wanted to determine if this regulation is restricted to high risk viruses or whether low risk viruses also upregulate SR protein expression. To this end I transiently transfected GFP expression constructs containing HPV16, HPV11 or HPV6 E2 into 293-T cells and monitored expression of SRSF3. GFP HPV E2 plasmids were a gift of Dr Kevin Gaston (University of Bristol). SRSF3 was chosen as 293-T cells express high levels of SRSF1 naturally and this may mask any differences in expression. I did not examine SRSF2 because antibodies against this protein do not work well in western blotting. Previously we have used the Mab104 that recognises all classical SR proteins. However the number of bands on the blot and the fact that the bands corresponding to SRSF1 and SRSF2 migrate very close to one another made differences in protein levels very difficult to determine. E2 constructs were transiently transfected in to 293-T cells using Lipofectamine reagent (Invitrogen) and incubated for 48 hours. After 48 hours protein was extracted into NP40 lysis buffer and the protein concentration determined by Bradfords Assay. Equal concentrations of protein were loaded in each well and fractionated on a 4-12% denaturing SDS-PAGE in SDS buffer. The proteins were transferred onto a nitrocellulose membrane and western blotted for SR protein expression. The results from one experiment are shown in Figure 3.8. γ-tubulin is used as a loading control standard to demonstrate equal concentrations of

protein are being compared. The antibodies being detected are indicated to the right of the gel. Track 1 shows 293-T cells alone with no transfection. Track 2 contains extracts from 293-T cells treated with Lipofectamine alone, this control was included as Lipofectamine can be toxic to cells and I wanted to ensure that the Lipofectamine was not altering protein expression through cellular cytotoxicity. Track 3 contains protein extracts from 293-T cells transfected with an empty GFP vector and is an appropriate control for this experiment. Tracks 4 and 5 contain HPV6 E2 transiently transfected 293-T cells with 100ng or 500ng of E2 respectively. Two concentrations of E2 were tested as E2 can cause cell death and can also function as a repressor at high concentrations therefore two concentrations were used to ensure optimal activation of SRSF2. In track 5 there is a small increase in SRSF3 expression with the higher concentration of E2 suggesting HPV6 E2 may be upregulating SRSF3 expression. Tracks 6 and 7 contain protein extracts from 293-T cells that were transiently transfected with HPV11 E2 at 100ng or 500ng respectively. In these extracts there is a clear increase in SRSF3 expression when compared to the empty GFP vector alone (track 3). Similarly in tracks 8 and 9 where HPV16 E2 has been transiently transfected there is an increase in SRSF3 levels as has been reported previously (Mole et al., 2009a, Mole et al., 2009b). The experiment was carried out three times with similar results. HPV16 has been shown to bind to SP1 and transactivate the promoter of SRSF1 (Mole et al., 2009b). Similarly, HPV11 E2 has SP1 binding sites, which suggests that HPV11 E2 could have similar transactivation functions as HPV16 E2 through the interaction with SP1. These data suggest that low risk viruses may also regulate SR protein expression which suggests that these SR proteins are essential for a productive viral life cycle.

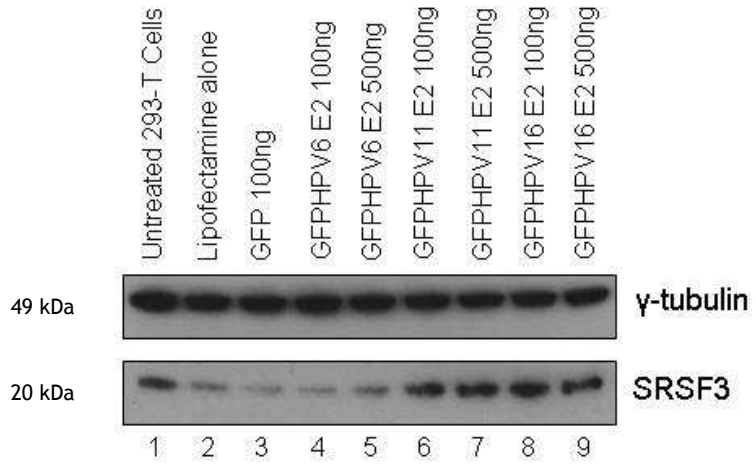


Figure 3.8 SRSF3 is upregulated by high and low risk papillomaviruses. Western Blot analysis of SRSF3 protein levels in transiently transfected 293-T cells. The transfected expression constructs are indicated to the top of the blots. **A)** Blot probed with GTU88 against γ-tubulin and used as an internal control. **B)** Blot probed with Mab7B4 showing SRSF3 levels in the transfected cells.

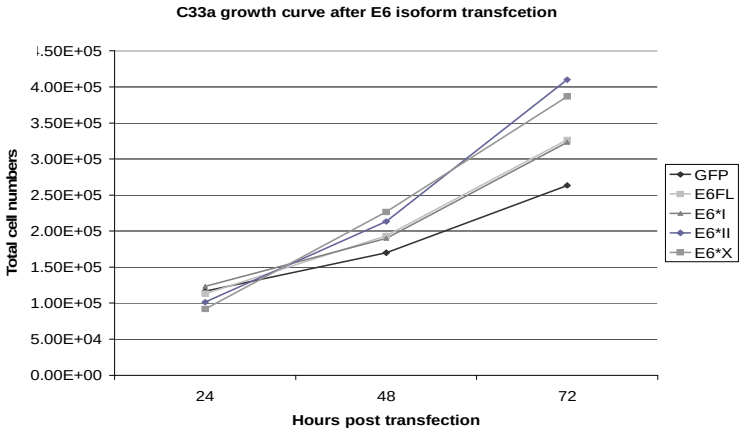
It is not surprising that low risk viruses also require SR protein expression as SR proteins are expected to be required for alternative splicing and other RNA processing functions in the cells.

In conclusion high risk viruses are not alone in their capacity to upregulate expression of SR proteins, specifically SRSF3. Therefore treatment of both high risk and low risk viruses could possibly be carried out by targeting of the same SR protein.

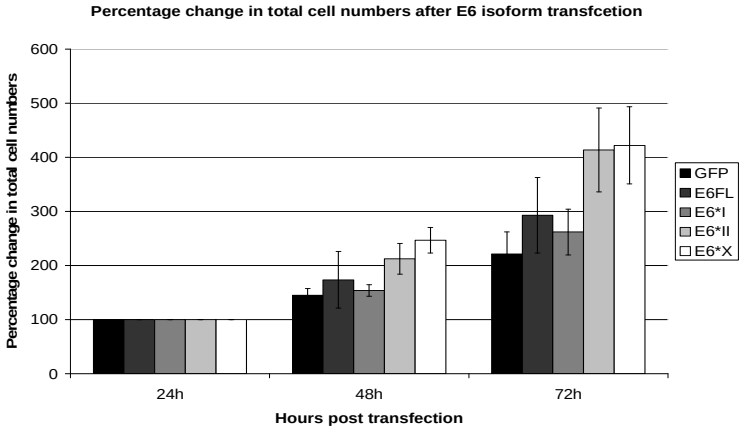
3.1.6 Small E6 isoforms promote cell growth

The final aim for this chapter was to identify any function for the small HPV16 E6/E7 isoforms that become up regulated upon transformation of the W12 epithelial cells. As these isoforms are upregulated upon transformation of the epithelial cells it is likely they will have some function that is associated with tumourigenesis. To test any oncogenic function, I created constructs containing cDNAs encoding each E6/E7 isoform cloned into the GFP expression vector pEGFP-C1 (Clontech). A full length E6 construct cloned into p3XFLAG vector (Sigma) was also used in this experiment (created by Dr Alasdair MacDonald). The constructs were transfected individually into C33a HPV-negative cervical epithelial cells using Lipofectamine (Invitrogen) and the transfection efficiency monitored by GFP expression. The transfection efficiency seemed similar in all cases at around 70%. The number of cells was counted after 24, 48 and 72 hours of transfection for each experiment. Figure 3.9 A shows the result from the growth curve study. Compared to cells transfected with a GFP-expression

A


C33a growth curve after E6 isoform transfcetion
Total cell numbers
1.50E+05
4.00E+05
3.50E+05
3.00E+05
2.50E+05
2.00E+05
1.50E+05
1.00E+05
5.00E+04
0.00E+00
24
48
72
Hours post transfection
GFP
E6FL
E6*I
E6*II
E6*X

B


Percentage change in total cell numbers after E6 isoform transfcetion
Percentage change in total cell numbers
600
500
400
300
200
100
0
24h
48h
72h
Hours post transfection
GFP
E6FL
E6*I
E6*II
E6*X

C

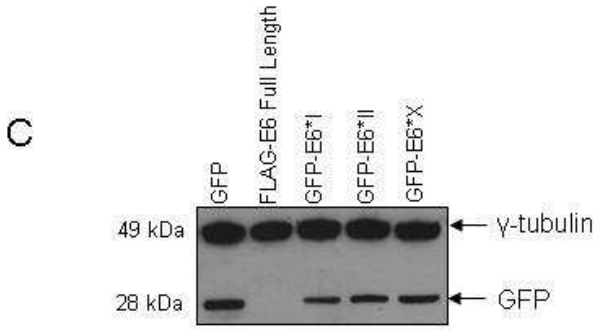
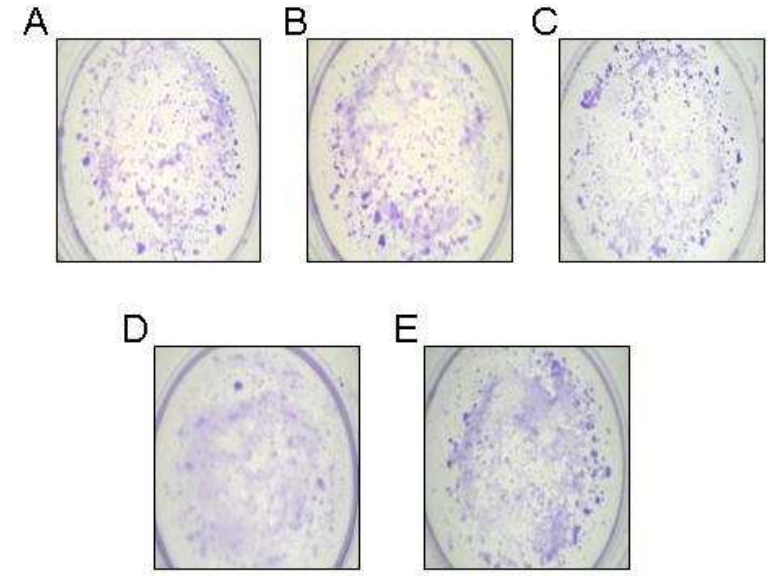

GFP
FLAG-E6 Full Length
GFP-E6*I
GFP-E6*II
GFP-E6*X
49 kDa
28 kDa
γ-tubulin
GFP

Figure 3.9 Transient transfection of E6 isoforms promotes cell growth. A) Growth curve analysis of the total C33a cell numbers after 24, 48 and 72 hours of transfection with expression constructs for each of the E6 isoforms. The isoform that was transfected is indicated to the right of the graph. One experiment is shown in the graph. The experiment was carried out three times. **B)** The percentage change in total cell numbers is shown in the bar chart. Cell numbers at 24 hours were set at 100%. The results of three independent experiments are shown in the chart. The data show the mean and standard deviation in each case. The transfected isoforms are indicated to the right of the graph. **C)** Western blot analysis of E6 construct expression levels. Blot probed with Mab GTU88 against γ-tubulin and Mab 2555 against GFP. The protein being detected is indicated to the right of the blot.

construct control, C33a cells transfected with constructs expression E6FL or E6*I grew faster and each grew at a similar rate. In contrast, cells transfected with expression constructs for E6*II and E6*X grew even faster with an almost 2-fold increase in cell numbers relative to control transfected cells (Figure 3.9 A) The graphs indicate that transfection of the cells with the smaller isoforms, E6*II and E6*X promotes cellular proliferation. This finding could explain why there are increased levels of these isoforms in transformed epithelial cells only. It is also interesting that in these experiments E6*I seems to have a negative effect on cell growth as the total cell numbers are decreased (Figure 3.9 B 72 hour time point) when compared to the cells expressing full length E6/E7. This however seems to fit with other reports of E6*I having antagonistic functions to full length E6. Cells expressing GFP-tagged constructs appeared to be transfected to similar levels, with around 70-80% of cells being transfected. Therefore differences in growth are probably not due to differences in transfection efficiences or expression levels. Similarly, GFP protein levels are similar.

Anchorage independent growth assays were also performed using the same E6 isoform constructs transfected into the HPV-negative cervical epithelial tumour cell line, C33a. Cells were transfected with the E6 isoform plasmids and grown on soft agar for 14 days as described in section 2.2.2.8. The top layer of media was removed and the colonies stained with 0.005% crystal violet for 1 hour. The cells were then dried and photographed and one experiment shown in Figure 3.10. No reproducible, statistically significant, difference was seen between the control cells and any of the transfected isoforms. The reason for this could be that there is a combinatorial effect of the isoforms when it comes to transforming the cells and no single isoform is solely responsible for the induction of anchorage independent growth.



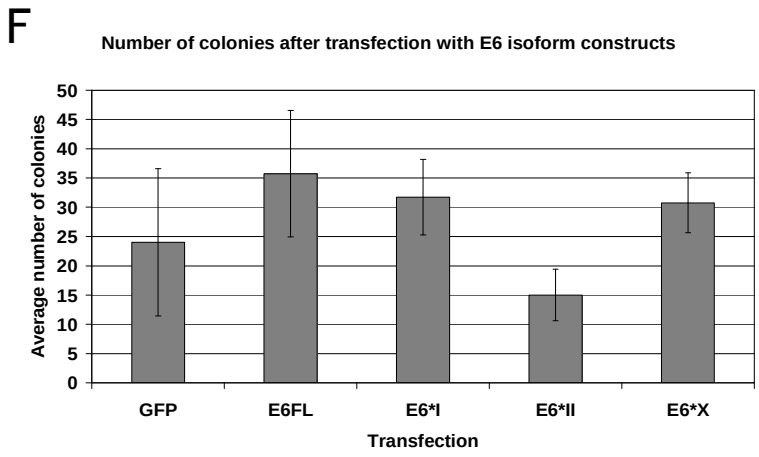


Figure 3.10 Colony formation assay after transient transfection of E6 isoforms. **A)** Colonies formed after transient transfection with empty GFP vector. **B)** Colonies formed after transient transfection with a full length E6/E7 vector. **C)** Colonies formed after transient transfection with an E6*I expressing plasmid. **D)** Colonies formed after transient transfection with an E6*II expressing vector. **E)** Colonies formed after transient transfection with a plasmid expressing E6*X. Each picture is from a single experiment. **F)** Quantification of the number of colonies present after each transfection from three independent experiments.

3.2 Discussion

HPV16 E6/7 mRNAs can be alternatively spliced, with at least four possible mRNA isoforms being produced. It was originally thought that the alternative splicing of E6 allowed for efficient translation of E7 (Sedman et al., 1991). However, a number of studies over many years have demonstrated that translation of E7 is just as efficient from full length and spliced E6/E7 mRNAs. These studies were carried out by overexpression of HPV16E6E7 constructs. Therefore in these analyses E6 and E7 were not under the control of the viral promoter, but under a strong alternative promoter such as CMV. Also the polyadenylation site provided in the plasmid construct is a strong SV40 polyadenylation site and does not reflect the efficiency of any of the viral polyadenylation sites. Both these facts mean that splice site selection and translation efficiency for E6 and E7 may be altered. When proteins are overexpressed in cells they will not have the same biological activities as the protein as expressed normally because there may be saturation of protein-protein interactions or an induction of novel interactions that do not occur with normal protein levels. Moreover, the translation apparatus may be overwhelmed with mRNAs affecting translation efficiency. Therefore it is not possible to say whether E7 is translated more efficiently or not from spliced E6 transcripts.

Alternative splicing of E6/E7 mRNAs in our W12E cell line does not appear to be altered upon differentiation of the infected epithelial cells. Even when the total E6/E7 RNA levels increased in differentiated W12E cells (Figure 3.2 C, track 5) compared to undifferentiated W12E cells (Figure 3.2 C, track 3), the isoform ratios did not appear to change. That does not mean, however that the splicing of E6/E7 is not important for the viral life cycle. As has been suggested by a number of studies it is perhaps the ratio between the different E6 isoforms that

is important for determining the function of E6 at any one stage in the life cycle. Like other studies, this work has shown that E6*I is the predominant isoform expressed in HPV-infected cells regardless of differentiation or transformation status. Lack of any apparent change in isoform expression upon differentiation of W12 epithelial cells could suggest that there is tight control over expression levels of each isoform in relation to the levels of the other isoforms. Again this would point towards the ratios of the isoforms being important for E6 function during the virus life cycle. Alternatively, the lack of any qRT-PCR data means that I cannot rule out subtle changes in E6 isoform expression not revealed by semi-quantitative RT-PCR.

Transformation of the epithelial cells on the other hand seems to have a significant effect on E6 splicing. Transformation seems to increase expression of the smallest two isoforms, E6*II and E6*X. The smallest isoform E6*X is only detectable, in these assays, in fully transformed W12GPXY and CaSki cells (Figure 3.3 C tracks 7 and 9 and Figure 3.4 C tracks 5 and 7). Altered isoform expression and altered expression levels during tumour progression could explain why during transformation the E6 protein interactions essential for viral replication become oncogenic. If the levels of E6 protein are the key to the function of E6 then there is likely to be a strict control over the ratio of expression between the isoforms. However when the proteins become overexpressed, as they are during transformation, then such strict control may be lost. If control over isoform expression is lost then loss or changes in patterns of interactions may aid in the transformation of the cells. It is not likely that the E6*X isoform is only expressed in these cells but that the expression is increased to detectable levels. In fact, Milligan et al. (2007) identified at least one polycistronic mRNA expressing E6*X in their analysis of transcripts isolated

from differentiated W12E cells encoding late HPV16 RNAs. These data could indicate that the smaller isoforms have functions, that when over expressed, have transforming effects on the infected cell. For example, they could be promoting cell growth and division. E6*II isoform also appears to be increased in levels upon transformation whereas E6*I is the dominant isoform in both non-transformed and transformed cells and level of the RNA do not appear to increase. It has been suggested that detection of high levels of E6*I could be indicative of high grade lesions and could be used as a diagnostic tool (Kosel et al., 2007). However my data suggests that this might not be a useful assay because E6*I is highly expressed even in W12E cells where there is a productive viral infection taking place. The data shown in Figures 3.2 and 3.3 require to be replicated in patient tissue to determine whether E6*II or E6*X are more readily detected in high grade lesions than in low grade lesions. If so, detection of these isoforms might be more useful in diagnostics.

If E6*X is in fact a tumour-specific isoform then it could also provide a possible therapeutic target for treatment of HPV16-positive cervical cancers. Cervical cancers are considered to be one of the more accessible tumours and therefore topical therapies could be designed to treat the tumours. This therapeutic avenue however, could only be used for HPV16-positive tumours as HPV18 does not produce E6*II or E6*X isoforms. Therefore therapy would only be effective in 40-50% of HPV-positive cervical cancers. It is interesting that HPV18 does not express E6*II or E6*X isoforms, in my opinion this may explain why HPV16-infections are more commonly linked to squamous cell carcinoma and HPV18 infections are more commonly linked to adenocarcinomas (Bulk et al., 2006). This suggests that the smaller isoforms may have a role infection of the

squamous epithelial cells in the ectocervix but are not necessary for infection of the epithelial cells in the endocervix.

Alternative splicing requires SR splicing regulatory proteins to identify exonic splicing enhancer elements and define exon-intron boundaries. Due to the alteration in E6/E7 mRNA splicing upon transformation of the epithelial cells, the levels of individual SR proteins were studied. SR proteins 1, 2 and 3 are all upregulated upon transformation of the cells (Figure 3.5), and in patient tissues (Figure 3.6) (Mole et al., 2009a). This could be why E6/E7 RNA splicing is altered during transformation. Overexpression of an individual or a combination of SR proteins may be changing the splice site usage and thereby disrupting the ratios of E6 isoforms. If one of these upregulated SR proteins is responsible for enhancing the usage of the E6*X splice acceptor site at nt 742 then this SR protein could perhaps be targeted by small molecule inhibitors to remove the E6*X isoform from transformed cells thus also removing any tumour specific functions the isoform may have. Alternatively, siRNA targeted to the E6*X-specific splice junction would ablate this transcript isoform.

Papillomaviruses have a unique problem where viral replication and gene expression takes place in differentiating epithelial cells where SR protein expression is normally switched off. Therefore in order to express their RNAs and proteins, HPVs must control the expression of SR proteins. The analysis of SR protein levels in the W12 transformation model provides evidence that SRSF2 may be required for the completion of the viral life cycle. Previously in our laboratory we discovered that SRSF1 levels are upregulated in W12E cells. This is particularly striking when compared to the lack of upregulation in HaCaT or W12G cells (McPhillips et al., 2004). Mole et al. (2009) found a similar increase

in SRSF2 and 3 in W12E cells but not in HaCaT cells implying that virus infection was responsible for the change in SR protein levels. Here I have also demonstrated that the level of SRSF2 and SRSF3 are upregulated in W12E cells when compared with W12G cells. Preliminary evidence has suggested HPV16 E2 transcription factor upregulation of SRSFs 2 and 3 as well as SRSF1 (Mole et al., 2009b). Changes in levels of these other SR proteins may be relevant for the virus life cycle because unpublished data from our laboratory show that knock down of SRSF3 leads to a significant decrease in viral late mRNAs and late protein expression in differentiated W12E cells and in NIKS/HPV16 cells and previously SRSF3 was shown to be required to process late viral RNAs in U2OS osteosarcoma cells, although these tumour cells overexpress SR proteins (Mole et al., 2009b), and E6/E7 RNAs in CaSki cells (Jia et al., 2009).

From these data it is possible to suggest that SRSF2 may also have a role in processing virus RNAs during the life cycle the mechanism of which has yet to be determined. The viral regulation of these SR proteins underlines the importance of splicing to the completion of the viral life cycle and emphasises the complexity of the viral RNA processing regulation.

The overexpression of SR proteins is not unique to cervical tumour cells; in fact several SR proteins have been shown to be overexpressed in a number of cancers. However, the mechanism controlling this overexpression is not known in every case. For SRSF1, in breast cancers, it appears that the gene becomes duplicated during transformation (Karni et al., 2007). Now I have shown that SRSF1 gene copy number is also increased in cervical tumour cells (Figure 3.7) making gene amplification a likely mechanism for SRSF1 overexpression in

cancers in general. The 17q23 has been shown to be amplified in a number of cancers including breast, testicular, ovarian and lung tumours (Kallioniemi et al., 1994, Ried et al., 1994, Korn et al., 1996, Hirasawa et al., 2003). Gene amplification explains why there is more SRSF1 expressed in cancer cells than normal cells. In contrast, SRSF2 and SRSF3, in the W12 cervical tumour cells (Figure 3.7), do not seem to be overexpressed due to gene amplification. It is possible that their overexpression is carried out at the transcriptional level or through an alteration to the protein stability. These mechanisms may also regulate SRSF1 together with gene amplification. These data along with other evidence from our laboratory are the first demonstration of SR protein overexpression in cervical cancer. SRSF1 and SRSF3 have been shown to be oncogenic in breast and ovarian cancers as reduction in protein levels can revert the tumour phenotype (Karni et al., 2007, He et al., 2011). From these data it is reasonable to suggest that SRSF2 may also be considered to be an oncoprotein.

The overexpression of the SR proteins will most likely, not only effect the viral gene expression prolife, but the cellular gene expression profile (reviewed by (Ward & Cooper, 2009). SR proteins could then provide an attractive therapeutic target for not only cervical but other cancers as well.

4 Results 2

4.1 Introduction

Results in Chapter 3 demonstrated alternative splicing of HPV16 E6/E7 is altered upon transformation of W12 cervical epithelial cells. Cellular transformation also caused a change in the levels of certain SR proteins; namely SRSF1, 2 and 3. It is possible therefore that the increase in SR protein levels in the fully transformed cervical epithelial cells is responsible for the change in splicing of E6/E7 mRNA. SR proteins are cellular splicing regulators and their most well known function is to define exon/intron boundaries and promote RNA splicing. Alternative splicing is a way in which one RNA molecule can encode multiple proteins with differing functions. SR proteins bind to RNA and direct which exons and introns are to be included or excluded from the resulting mRNA. Control of RNA splicing is tightly regulated during development and differentiation with numerous proteins involved. Indeed, aberrant splicing is associated with many human diseases. SR proteins recruit spliceosomal components to the pre-mRNA by binding to enhancer elements within the mRNA. SR proteins help to define exons by facilitating interactions between U1 snRNP bound at the 5' splice site and the 65 kDa subunit of U2AF which is bound to the 3' splice site (Robberson et al., 1990). Similarly, SR protein interactions with U1 snRNP and U2AF35 can also bridge interactions across introns (Wu & Maniatis, 1993). Finally SR proteins (SRSF1 in particular) can bind at the branch point and facilitate spliceosome assembly (Shen et al., 2004). All of these steps are tightly regulated and levels of SR proteins are known to contribute to this regulation.

Therefore increased expression of any SR protein may lead to the splicing process becoming deregulated and could be catastrophic for the cell.

In addition to well documented roles in splicing and its regulation, SR proteins have other RNA-related functions in the cell. For example, SR proteins have been shown to be involved in regulating RNA stability, RNA export from the nucleus and translation of the RNA (Huang & Steitz, 2001, Lemaire et al., 2002, Huang et al., 2003, Huang et al., 2004, Sanford et al., 2004, Zhang & Krainer, 2004, Michlewski et al., 2008). This means that there are a number of RNA processing steps where increased levels of SR proteins either during HPV-infected cervical epithelial differentiation or HPV-associated cervical tumour progression, could be acting upon E6/E7 expression.

SR proteins control accurate RNA processing which is essential for production of the normal cellular proteome. It is accepted that altered SR protein expression could lead to cell transformation. Overexpression could cause novel RNA splice isoforms to be produced. These could encode an oncogenic isoform of the wild type protein or a dominant-negative isoform that abrogated function of for example, a wild type tumour suppressor. Such possibilities have already been demonstrated by Karni et al. (2007). In this study SRSF1 overexpression was shown to be sufficient to induce anchorage independent growth and to immortalise rodent fibroblasts and NIH 3T3 cells. Injection of NIH 3T3 cells stably overexpressing SRSF1 subcutaneously into nude mice formed large sarcomas (Karni et al., 2007). Overexpression of SRSF1 protected the transfected cells from serum starvation-induced apoptosis and when co-transfected with H-Ras enhanced cellular proliferation (Karni et al., 2007). The study also showed that knockdown of SRSF1 overexpression, to levels that are

normally present in human lung cancer cells, was sufficient to reverse the transformed phenotype by reversing the cell's ability to form colonies in soft agar (Karni et al., 2007).

Significantly, while knockdown of SRSF1 altered the isoform expression of some of the cancer-related genes studied, it did not seem to affect expression of several endogenous genes that were also analysed (Karni et al., 2007). The study focused on several proto-oncogenes and tumour suppressors which are known to be alternatively spliced. Some targets, for example BIN1 which normally acts as a tumour suppressor, suppressing c-Myc activity, were found to have an increased level of one isoform upon SRSF1 overexpression that prevents the suppression of Myc activity (Karni et al., 2007). In contrast other targets such as Caspase 9 were not affected by SRSF1 overexpression. This is probably due to the functionally redundant nature of the SR proteins. SR proteins are structurally related proteins and bind to loose consensus sequences in the elongating mRNA strand. In a couple of studies knockdown of a single SR protein was not lethal to the cells and this is thought to be because other SR proteins can carry out the functions that would normally be carried out by the knocked down SR protein (Bakkour et al., 2007, Keriel et al., 2009).

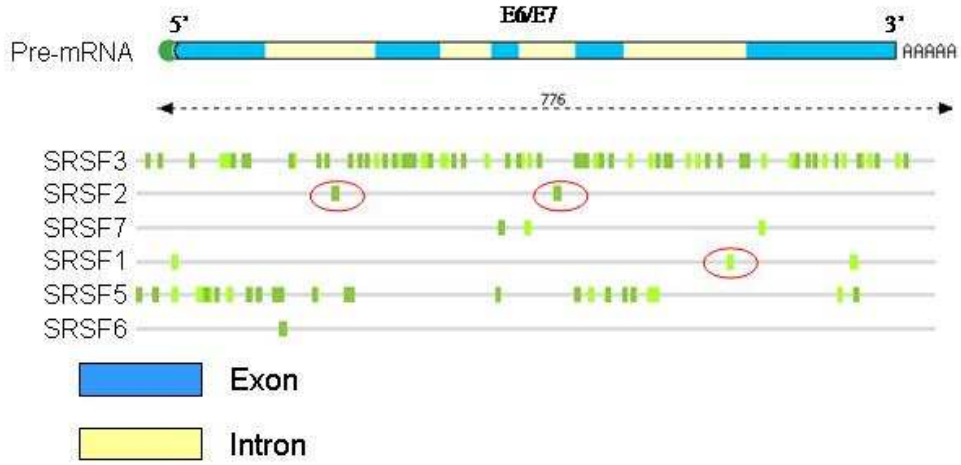
Very recently SRSF3 was identified as being oncogenic in ovarian epithelial cells (He et al., 2011). Knockdown of SRSF3 resulted in reduced cell growth and reduced the cells' ability to form colonies in soft agar by around 90%. Knockdown of SRSF3 also induced apoptosis in ovarian tumour cells. SRSF3 was knocked down using a doxycycline inducible siRNA for 5 days and the apoptotic cells detected by Hoechst 33342 staining. Hoechst 33342 stain binds to AT base pairs in DNA, however it stains condensed chromatin brighter than normal chromatin

giving an indication of apoptosis by counting the cells with condensed chromatin. There was a 40% increase in the number of cells positive for bright Hoechst stain when SRSF3 was knocked down compared to control. The authors then went on to implicate the intrinsic apoptosis pathway as being activated in response to SRSF3 knockdown. Cleaved caspase-9 was upregulated in the siRNA knockdown cells indicating that the intrinsic apoptosis pathway is active. From these data it could be concluded that SRSF3 knockdown reduces cell growth and colony formation and also results in apoptosis, which suggests that SRSF3 in these A2780 ovarian cancer cells may be oncogenic (He et al., 2011).

As these splicing factors, SRSF1 and SRSF3, have been implicated in enhancing tumour properties and I have shown them to be overexpressed in transformed cervical epithelial cells (Chapter 3 section 3.1.3), it is possible that they may be having a similar tumour promoting effect in cervical tumours. Similarly the reported altered splicing effects on a number of targets after SRSF1 overexpression prompted me to investigate whether the alteration to E6 splicing observed in the transformed cervical epithelial cells (Chapter 3 section 3.1.2) may be due to the overexpression of an SR protein. With this in mind there were two hypotheses I addressed in this chapter. Firstly, that the overexpression of one of the SR proteins is responsible for the upregulation of the small E6 isoforms in transformed cervical epithelial cells. Secondly, that the overexpression of an SR protein is contributing to the transformed phenotype of the epithelial cells.

4.1.1 SRSF2 controls HPV16 E6/E7 RNA expression in transformed cervical epithelial cells

As shown in Chapter 3, section 1.1.3 a subset of SR proteins are overexpressed in transformed cervical epithelial cells. In the same cells, E6/E7 RNA splicing pattern is altered. So I hypothesise that one or all of the overexpressed SR proteins is responsible for the change in viral RNA splicing. More specifically, the overexpression of an SR protein is responsible for the upregulation of E6*X expression. In order to determine whether any of the classical SR proteins could theoretically bind to the HPV16 E6/E7 mRNAs a bioinformatic analysis of the E6/E7 open reading frame was carried out. The full E6/E7 open reading frame sequence was entered into the Reg RNA sequence prediction website (http://regrna.mbc.nctu.edu.tw/index.php). The program analyses input sequences for consensus SR protein binding sequences. The resulting output is shown in Figure 4.1. Binding sequences are indicated by the presence of a vertical green bar, a green bar within 100 nucleotides of an exon junction is normally considered to be specific and functional. Therefore SRSF1, SRSF2 and SRSF7 could all be involved in processing of the E6/E7 mRNAs. SRSF3 and SRSF5 binding is not specific therefore they are unlikely to be involved in E6/E7 mRNA processing.



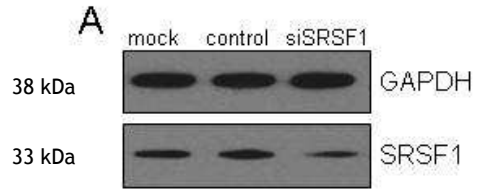
SRSF2 binding sequence: GRYcSYR (R, purine; Y, pyrimidine; S, G or C)

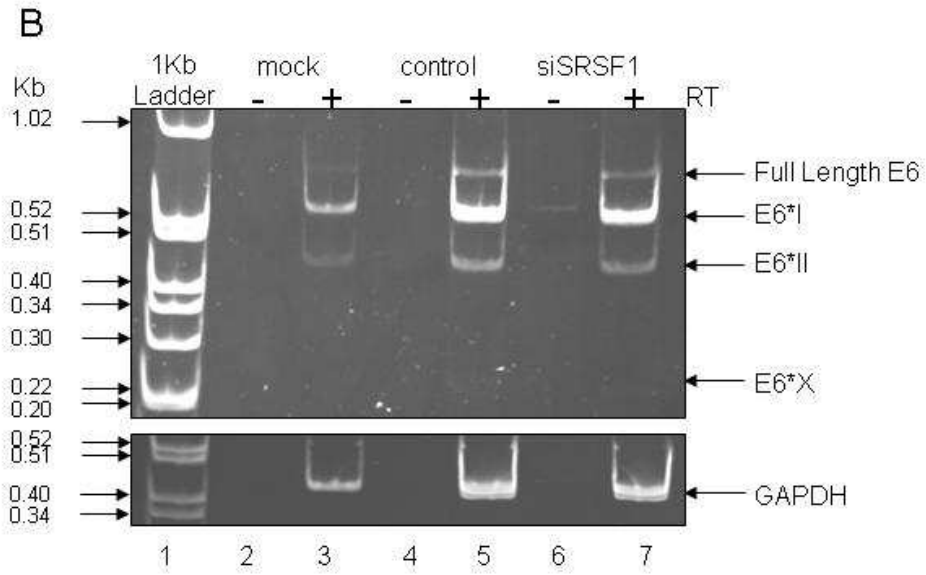
Possible HPV16 E6 binding sequence: GTATATAG or GCTCAGAG

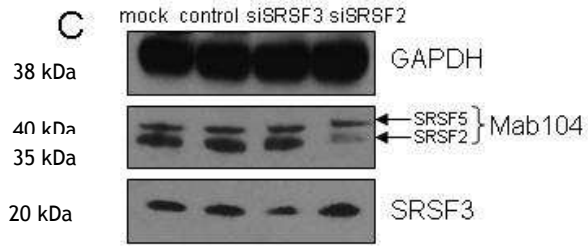
Figure 4.1 Computer prediction of SR protein binding sites in the HPV16 E6/E7 open reading frame. Online program into which the target sequence is entered and algorithm identifies possible binding sites for RNA binding proteins. Any binding sites within 100 nucleotides of a splice junction could be significant. Proteins with multiple binding sites such as SRSF3 are unlikely to be significant. http://regrna.mbc.nctu.edu.tw/index.php. Red circles identify those SR proteins which are predicted to specifically bind the target mRNA close to splice donor or acceptor sites.

In order to determine which, if any, SR protein contributed to E6/E7 RNA processing, the SR proteins SRSF 1-3 were individually knocked down using siRNA transfection in W12GPXY cells, the most transformed cell line displaying the highest levels of SR protein expression. SRSFs1-3 were chosen as they are overexpressed in transformed cervical epithelial cells and SRSF1 and 2 have been predicted to bind the E6/E7 RNAs. Although SRSF6 binding appears to be specific and unique the expression of the protein is not changed upon transformation, therefore I thought it unlikey to be involved in the alteration of E6 splicing upon epithelial transformation. All siRNA reactions were carried out using Lipofectamine RNAiMAX (Invitrogen). This reagent is formulated to specifically deliver siRNA molecules into cells with minimal toxicity. Total RNA was prepared followed by reverse transcription before being PCR-amplified using the same E6/E7 primers used in section 3.1.1 and listed in Table 1. All RNAs were DNase-1 treated before being reverse transcribed and amplified in a semi-quantitative PCR reaction using GAPDH primers as an internal control standard. The PCR products were resolved on a 6% acrylamide gel and stained with ethidium bromide. The results of this experiment are shown in Figure 4.2. Protein was also extracted from siRNA-treated cells and western blotted as described in section 2.2.3.2 to check for efficiency of knockdown of the various SR proteins. Knockdown of all three SR proteins was consistently achieved (Figures 4.2 A and C). Moreover, siRNA knockdown was specific because for example, SRSF2 knock down did not alter levels of SRSF3 and *vice versa* and SRSF5 was not altered by either knock down (Figure 4.2 C). However, knockdown was only around 50%. I considered that this was sufficient for the experiments, especially since SR proteins are essential cellular proteins required for splicing of cellular mRNAs and a higher percentage knockdown could compromise cellular functions. Figure 4.2 B shows the pattern of E6/E7 mRNA

expression after SRSF1 knockdown. Tracks 2, 4 and 6 show the reactions carried out in the absence of reverse transcriptase. No bands were detected confirming the absence of DNA from the initial RNA starting material. Track 3 shows the E6 isoforms expressed in W12GPXY cells without any treatment. Track 5 shows the E6 isoforms expressed after transfection with a fluorescent non-targeting siRNA called siGlo that was used to monitor transfection efficiency. The transfection efficiencies are detailed in Table 4. Track 7 shows the E6 isoforms expressed after SRSF1 knockdown. In all cases the E6 isoform pattern remained constant, with all isoforms being expressed. This result showed that SRSF1 is not required for E6/E7 isoform expression in W12GPXY cells. Figure 4.2 D shows the E6/E7 mRNAs expressed after SRSF2 and SRSF3 knockdown. Again tracks 2, 4, 6 and 8 all confirm the absence of DNA from the RNA starting material as the reactions have been carried out in the absence of reverse transcriptase. Tracks 3 and 5 show control experiments similar to those shown in tracks 3 and 5 of Figure 4.2 B. Track 7 shows the E6/E7 RNA isoform pattern expressed after SRSF3 knockdown. SRSF3 knockdown (Figure 4.2 C track 7) did not have any effect on E6/E7 mRNA expression as the isoform pattern and RNA levels are similar to those seen in tracks 3 and 5. Track 9 shows the E6/E7 RNA isoforms expressed after SRSF2 knockdown. Surprisingly, SRSF2 knockdown appeared to decrease total levels of all E6/E7 RNA isoforms. Although the levels of E6/E7 isoforms were very low it appeared that SRSF2 knockdown had no detectable effect on isoform pattern expression; isoforms E6 full length, E6*I and E6*II could still be observed with E6*I still being the most abundant isoform. These experiments were carried out three times with very similar results. Again quantification of the isoforms by qRT-PCR was not possible due to the difficulties in probe and primer design described in section 3.1.1. For each case in Figure 4.2 B and D


A
mock
control
siSRSF1
38 kDa
GAPDH
33 kDa
SRSF1


B
Kb
1Kb Ladder
mock
control
siSRSF1
- + - + - + RT
1.02
0.52
0.51
0.40
0.34
0.30
0.22
0.20
0.52
0.51
0.40
0.34
Full Length E6
E6*I
E6*II
E6*X
GAPDH
1 2 3 4 5 6 7


C
mock control siSRSF3 siSRSF2
38 kDa
GAPDH
40 kDa
35 kDa
SRSF5
SRSF2
Mab104
20 kDa
SRSF3

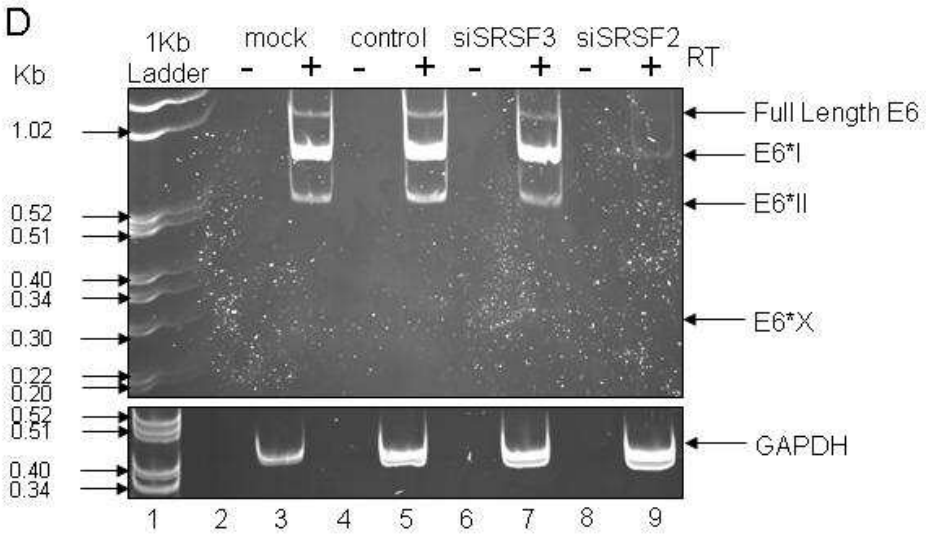
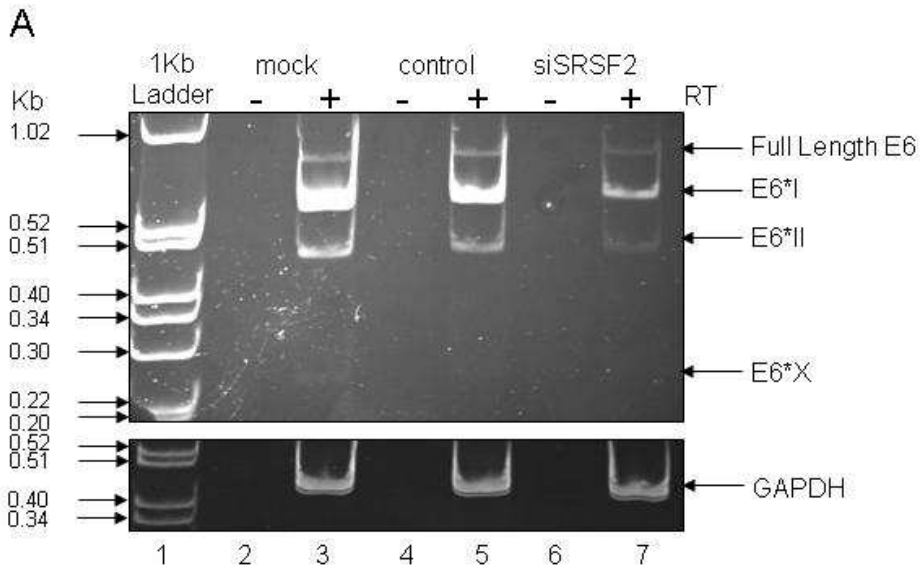

D
1Kb Ladder
mock
control
siSRSF3
siSRSF2
RT
Kb
1.02
0.52
0.51
0.40
0.34
0.30
0.22
0.20
0.52
0.51
0.40
0.34
Full Length E6
E6*I
E6*II
E6*X
GAPDH
1 2 3 4 5 6 7 8 9

Figure 4.2 E6 RNA expression is reduced upon SRSF2 knockdown in W12GPXY cells. A) Western blot showing knockdown efficiency of SRSF1 compared to control cells. Mab96 is used to detect SRSF1 levels and Mab6CS is used to detect GAPDH. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control. **B)** Upper panel Ethidium-bromide stained acrylamide gel electrophoresis of E6/E7 isoforms RT-PCR amplified from control and siRNA treated RNA from W12GPXY cells. Isoforms are indicated to the right of the gel. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "mock", untransfected RNA; "control", siGlo treated RNA; "siSRSF1", siSRSF1 treated RNA; Lower Panel Ethidium-bromide stained acrylamide gel electrophoresis of GAPDH RNA RT-PCR amplified from control and siRNA treated RNA from W12GPXY cells. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "mock", untransfected RNA; "control", siGlo treated RNA; "siSRSF1", siSRSF1 treated RNA; . **C)** Western blot showing knockdown efficiency of SRSF2 and SRSF3 compared to control cells. Mab7B4 is used to detect SRSF3 levels, Mab104 is used to detect SRSF2 levels and Mab6CS is used to detect GAPDH. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control. **D)** Upper panel Ethidium-bromide stained acrylamide gel electrophoresis of E6/E7 isoforms RT-PCR amplified from control and siRNA treated RNA from W12GPXY cells. Isoforms are indicated to the right of the gel. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "mock", untransfected RNA; "control", siGlo treated RNA; "siSRSF3", siSRSF3 treated RNA; "siSRSF2", siSRSF2 treated RNA; Lower Panel Ethidium-bromide stained acrylamide gel electrophoresis of GAPDH RNA RT-PCR

amplified from control and siRNA treated RNA from W12GPXY cells. RT, reverse transcriptase; “-“, RT-PCR reaction in the absence of reverse transcriptase; “+”, RT-PCR reaction in the presence of reverse transcriptase; “mock”, untransfected RNA; “control”, siGlo treated RNA; “siSRSF3”, siSRSF3 treated RNA; “siSRSF2”, siSRSF2 treated RNA.

GAPDH was used as a loading control and this control panel is shown beneath each RT-PCR gel. Similar to the isoform expression experiments in the previous chapter, it was important to test whether the effect of SRSF2 knockdown on E6/E7 RNA processing was unique to W12GPXY cells. Therefore a similar SRSF2 knockdown experiment was carried out in CaSki cells. Commercially available siRNA designed against SRSF2 (Dharmacon) was transfected into CaSki cells using Liopfectamine RNAiMAX (Invitrogen). RNA was extracted after transfection of the SRSF2 siRNA for 48 hours. All RNAs were DNase-1 treated before being reverse transcribed. cDNA was amplified using the same E6/E7 primers described in section 3.1.1. The cDNAs were amplified by semi-quantitative PCR using GAPDH primers as an internal control standard (Figure 4.3 A). The amplified products were resolved on a 6% acrylamide gel and post stained with ethidium bromide. Figure 4.3 A shows the amplified PCR products. Tracks 2, 4 and 6 show the reactions carried out in the absence of reverse transcriptase, thereby confirming the absence of DNA from the RNA starting material. Track 3 shows the E6/E7 isoforms expressed in untransfected CaSki cells. Track 5 shows the E6/E7 isoforms expressed after transfection of CaSki cells with a fluorescent non-targeting siRNA, siGLO. The transfection efficiency is listed in Table 4. Track 7 shows the E6/E7 isoforms expressed after SRSF2 knockdown. Significantly, SRSF2 knockdown in CaSki cells had a similar effect in reducing E6/E7 RNA levels, but not E6/E7 mRNA isoform expression pattern, as in W12GPXY cells (Figure 4.2 D, track 9). This suggests that SRSF2 may be responsible for processing of HPV16 E6/E7 RNAs in virus-infected transformed cells. As there is no SRSF2 commercially available that works well in Western blots, Mab104 was used to determine knockdown. However, as this antibody detects phosphorylated proteins it may also be true that it is the loss of


A
1Kb Ladder
mock
control
siSRSF2
RT
Kb
1.02
0.52
0.51
0.40
0.34
0.30
0.22
0.20
0.52
0.51
0.40
0.34
Full Length E6
E6*I
E6*II
E6*X
GAPDH
1 2 3 4 5 6 7

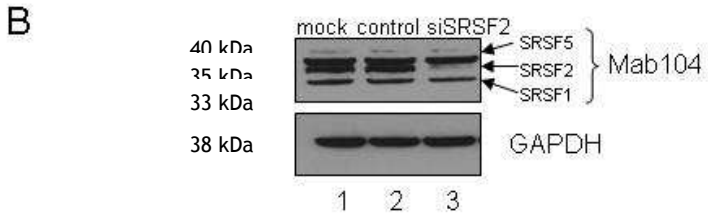
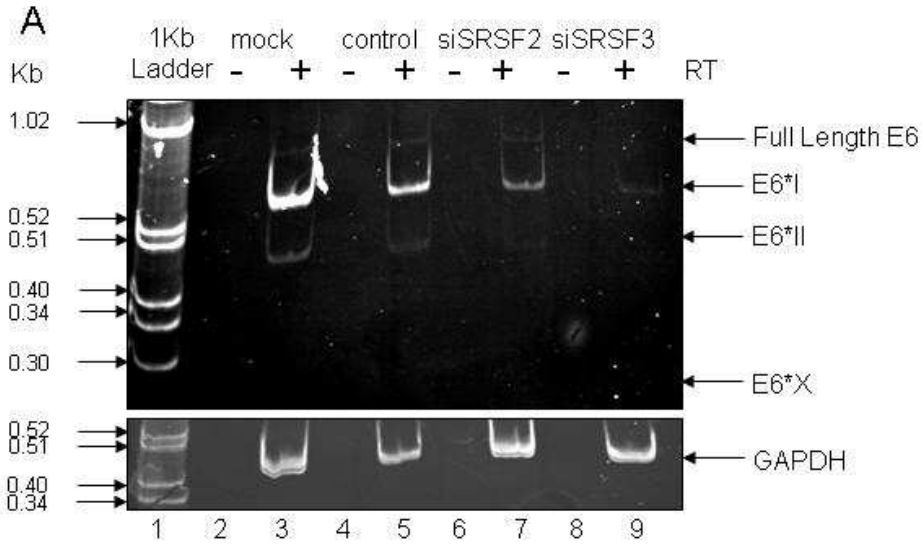

B
mock control siSRSF2
40 kDa
35 kDa
33 kDa
38 kDa
SRSF5
SRSF2
SRSF1
Mab104
GAPDH
1 2 3

Figure 4.3 E6 RNA expression is also reduced upon SRSF2 knockdown in CaSki cells. A) Upper panel Ethidium-bromide stained acrylamide gel electrophoresis of E6/E7 isoforms RT-PCR amplified from control and siRNA treated RNA from CaSki cells. Isoforms are indicated to the right of the gel. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "mock", untransfected RNA; "control", siGlo treated RNA; "siSRSF2", siSRSF2 treated RNA; Lower Panel Ethidium-bromide stained acrylamide gel electrophoresis of GAPDH RNA RT-PCR amplified from control and siRNA treated RNA from CaSki cells. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "mock", untransfected RNA; "control", siGlo treated RNA; "siSRSF2", siSRSF2 treated RNA. **B)** Western blot showing knockdown efficiency of SRSF2 and SRSF3 compared to control cells. Mab104 is used to detect SRSF2, Mab 7B4 is used to detect SRSF3 and Mab6CS is used to detect GAPDH. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control.

phosphorylation of SRSF2, and not a reduction in protein levels, that is reducing the E6/E7 RNA levels.

4.1.2 SRSF3 not SRSF1 or SRSF2 is required for E6/E7 RNA expression during the virus life cycle

My novel data suggests that SRSF2 but not SRSF1 or SRSF3 may be required for E6/E7 RNA processing in transformed cervical epithelial cells. In contrast with my results in W12GPXY cells, SRSF3 has previously been shown to be required for HPV16 E6/E7 expression in CaSki cells (Jia et al., 2009). I did not test knock down of SRSF3 in CaSki cells. However, I decided to investigate whether SRSF2 and/or SRSF3 could also control E6/E7 RNA expression during the viral life cycle. Compared to cervical cancer cells, there are only low levels of SRSF2 in W12E cells (Figure 3.4). W12E cells were grown on feeder layers for 4 days before being re-plated without feeder cells in keratinocyte growth medium (Lonza) without antibiotics. The cells were siRNA treated with siSRSF2 or siSRSF3 on day five and incubated for 72 hours before RNA was extracted. All RNA was DNase-1 treated before being reverse transcribed. The cDNA was amplified using the E6/E7 primers described in section 3.1.1. The amplified PCR products were resolved on a 6% acrylamide gel and stained with ethidium bromide. The PCR products are shown in Figure 4.4 A. In W12E cells SRSF2 knockdown had some effect on the levels of E6/E7 RNAs or mRNA isoforms. However in contrast to the results with W12GPXY and CaSki cells, SRSF3 knockdown resulted in very significantly reduced levels of E6/E7 RNAs. In summary, SRSF3 had no effect on expression of integrated E6/E7 open reading frames in W12GPXY cells but had a


A
1Kb Ladder
mock
control
siSRSF2
siSRSF3
RT
Kb
1.02
0.52
0.51
0.40
0.34
0.30
Full Length E6
E6*I
E6*II
E6*X
GAPDH
1 2 3 4 5 6 7 8 9

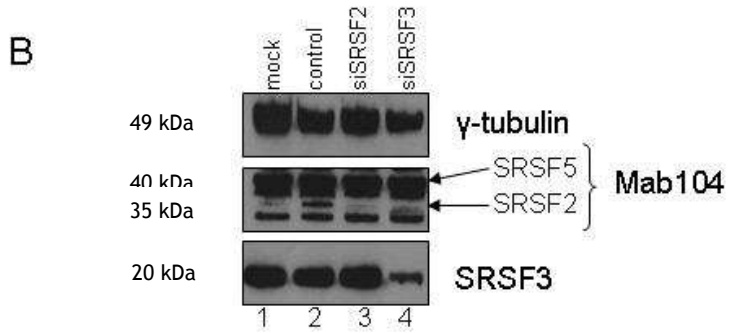

B
mock
control
siSRSF2
siSRSF3
49 kDa
γ-tubulin
40 kDa
35 kDa
SRSF5
SRSF2
Mab104
20 kDa
SRSF3
1 2 3 4

Figure 4.4 E6 RNA expression is reduced upon SRSF3 knockdown in W12E cells. A) Upper panel Ethidium-bromide stained acrylamide gel electrophoresis of E6/E7 isoforms RT-PCR amplified from control and siRNA treated RNA from W12E cells. Isoforms are indicated to the right of the gel. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "mock", untransfected RNA; "control", siGlo treated RNA; "siSRSF2", siSRSF2 treated RNA; "siSRSF3", siSRSF3 treated RNA; Lower Panel Ethidium-bromide stained acrylamide gel electrophoresis of GAPDH RNA RT-PCR amplified from control and siRNA treated RNA from W12E cells. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "mock", untransfected RNA; "control", siGlo treated RNA; "siSRSF2", siSRSF2 treated RNA; "siSRSF3", siSRSF3 treated RNA. **B)** Western blot showing knockdown efficiency of SRSF2 and SRSF3 compared to control cells. Mab104 is used to detect SRSF2 protein, Mab7B4 is used to detect SRSF3 protein and GTU88 is used to detect γ-tubulin protein. The protein being detected is indicated to the right of the blot. γ-tubulin is used as a loading control.

very significant effect on episomal E6/E7 expression in W12E cells. These differences are probably due to the fact that E6/E7 is expressed from a very different mRNA population in W12E cells (Milligan et al., 2007). This result also highlights the potential role in splicing regulation of the key differences I observed in SRSF protein levels in W12E and in the W12 transformed cell lines. Even subtle changes in SR protein levels in cells can result in quite different patterns of gene expression. It remains to be seen what other effects SRSF3 has on episomal HPV16 gene expression.

4.1.3 SRSF2 knockdown in transformed cells functionally impairs E6 protein

As has been shown in Figure 4.2 and 4.3, SRSF2 knockdown decreases E6/E7 mRNA levels in cervical cancer cells. However, changes in RNA expression may not always be reflected in an effect at the protein level due to regulation of RNA stability and translation efficiency. It is very difficult to study E6 protein levels due to low levels of expression, even in HPV-positive cervical cancer cells and due to a lack of suitable antibodies. E7 antibodies are better at detecting E7 protein in such cells. However, I decided to use surrogate markers of E6/E7 expression to examine the role of SRSF2 in the cervical tumour phenotype. I examined E6 and E7 separately and then concentrated on E6 function because three out of four of the E6/E7 mRNA isoforms are predicted to encode E6 isoforms and not E7. p53 is often used as an indication of E6 presence and function and pRb as an indication of E7 presence and function. Therefore SRSF2 was knocked down for 48 hours in W12GPXY cells using commercially available

siRNAs (Dharmacon). Protein was then extracted into NP40 lysis buffer containing protease and phosphatase inhibitors. The proteins were quantified by Bradfords protein assay and equal concentrations of proteins were denatured and resolved on a denaturing SDS-PAGE gradient gel in SDS running buffer. The proteins were then western blotted by transferring onto a nitrocellulose membrane and incubating with protein-specific antibodies. Detection of the protein was done using horseradish peroxidise tagged secondary antibodies followed by ECL detection. The results of the experiment are shown in Figure 4.5. Figure 4.5 A track 1 shows the protein levels in untransfected cells. Track 2 shows the protein levels in the cells after transfection with a non-E6/E7 targeting fluorescent siRNA (Dharmacon). Track 3 shows p53 protein levels after SRSF2 knockdown. The proteins being detected are indicated to the right of the membrane. In track 3 the p53 protein levels are increased when SRSF2 is knocked down corresponding to the observation that E6/E7 RNA levels decrease upon SRSF2 knock down (Figure 4.2 D). Figure 4.5 B shows the result of a quantification of three independent experiments. The increase in p53 levels was found to be significantly different, with a p value of 0.021, between control siRNA and SRSF2 knockdown samples after carrying out a student's T-test. This indicates that not only is the E6/E7 RNA reduced but oncoprotein levels are either correspondingly reduced or E6 protein is functionally impaired because increased p53 levels implies a reduction in E6-mediated p53 degradation.

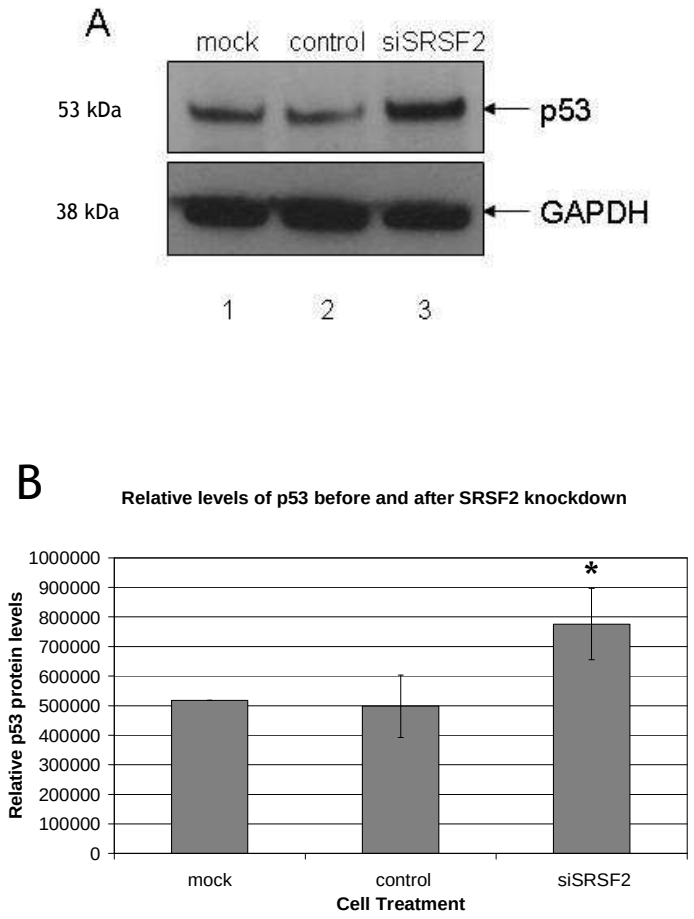


Figure 4.5 SRSF2 knockdown results in increased p53 levels. A) Western blot showing p53 levels in W12GPXY cells after SRSF2 knockdown. MabDO-7 is used to detect p53 protein and Mab6CS is used to detect GAPDH protein levels. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control. **B)** Graph showing the mean and standard deviation from the mean from three independent experiments. Pixel density was measured using Adobe photoshop. The increase in p53 levels after SRSF2 knockdown is statistically significant as indicated by "*".

A similar experiment was carried out using pRb as a surrogate marker to investigate levels of E7 protein. Figure 4.6 shows the protein levels of pRb with and without SRSF2 knockdown. The same protein extracts that were used in experiment 4.5 were western blotted for pRb levels. Track 1 shows the pRb levels in W12GPXY cells transfected with siGlo for 48 hours. Track 2 shows the pRb levels after SRSF2 knockdown for 48 hours. GAPDH is used as an internal loading control. The protein being detected is indicated to the right of the blot. Whereas in the previous experiment p53 levels increased upon SRSF2, pRb levels dropped (Track 2). This reduction in pRb protein levels was not highly significant as after a student's T-Test the p value was 0.08. It has been reported that pRb is phosphorylated however the antibody used detects total pRb levels and does not recognise phosphorylated pRb forms. This seemingly paradoxical finding is at first confusing as the expectation would be that if E6/E7 mRNAs are reduced upon SRSF2 knockdown and E6 protein is functionally impaired, then one would expect E7 to be likewise impaired. If E7 protein was reduced or impaired, according to the literature, pRb levels should increase. A possible explanation for this finding could be that the increase in p53 levels is causing the cell to switch off expression of cell cycle associated proteins and prepare to enter apoptosis. Due to the unexpected pRb finding further experiments were designed around the E6 impairment and the subsequent increase in p53.

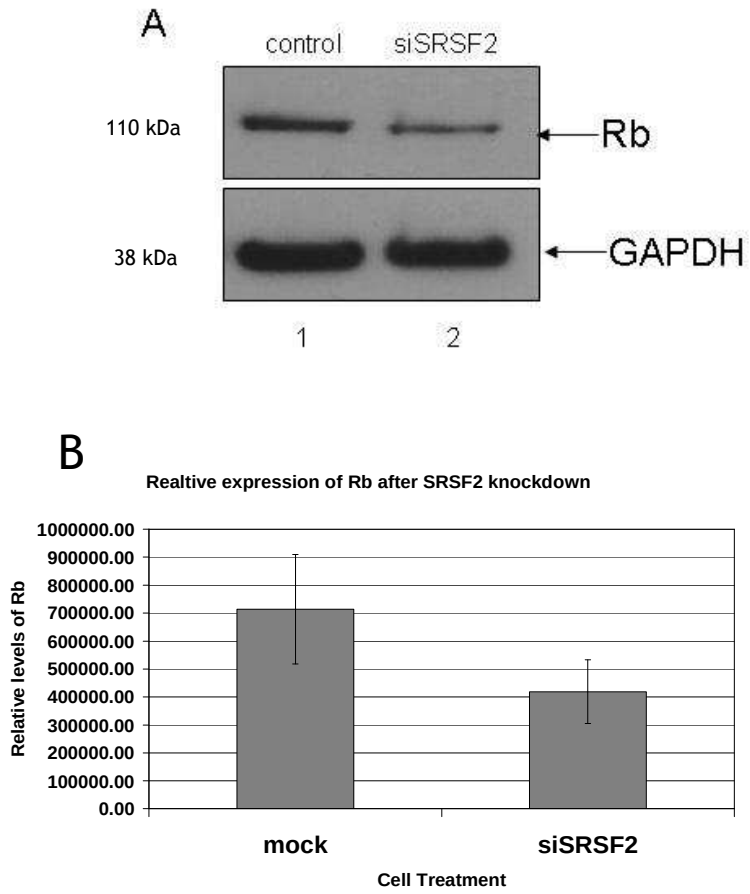


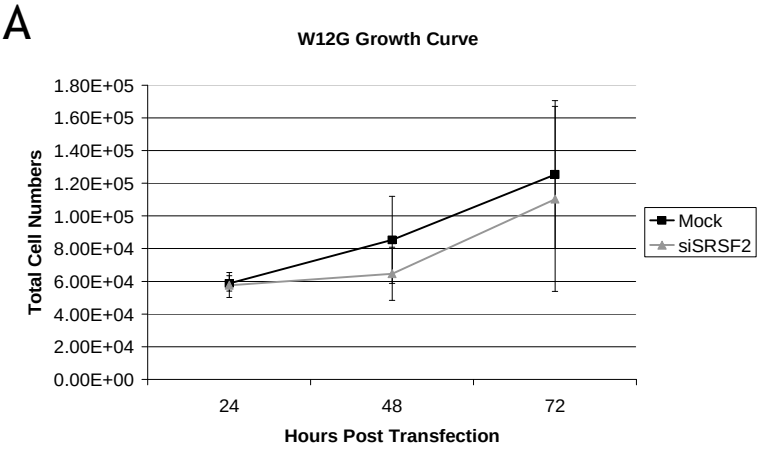
Figure 4.6 SRSF2 knockdown results in decreased pRb levels. A) Western blot showing Rb levels in W12GPXY cells after SRSF2 knockdown. Mab 4H1 is used to detect pRb and Mab6CS is used to detect GAPDH levels. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control. **B)** Graph showing the mean value and the standard deviation from the mean from three independent experiments. The pixel density was measured using Adobe photoshop.

4.1.4 SRSF2 contributes to the transformed phenotype of the cells

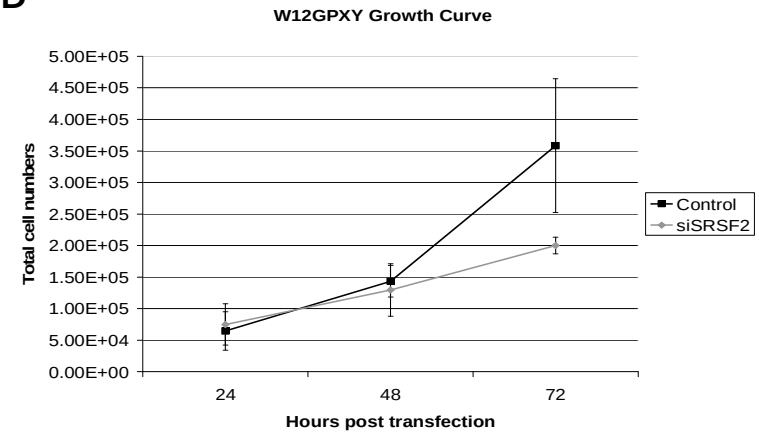
4.1.4.1 SRSF2 knockdown decreases cellular growth rate, alters cell morphology and interferes with the cell cycle

As E6 and E7 both cooperate to promote cell growth, the consequences of SRSF2 knockdown, and the subsequent reduction in E6/E7 RNA and by implication, E6 protein levels, on cell growth was investigated. W12G or W12GPXY cells were reverse transfected with commercially available SRSF2 siRNA (Dharmacon) and plated. W12G cells were grown on J2 3T3 feeder layer cells until day 5 when the feeder layer was removed and the cells counted, transfected and plated out. A reverse transfection was used in this instance instead of the usual forward transfection as I wanted to ensure that cell numbers were equal and to prevent the control cells from becoming confluent at the last time point. At 24, 48 and 72 hours of transfection, cells were harvested and counted. Figure 4.7 A and B show the growth curves of W12G and W12GPXY over the course of the experiment. Compared to mock-transfected cells, SRSF2 knockdown had no significant effect on cell growth rate in W12G cells (Figure 4.7 A). This result is as expected because these cells are not transformed and do not have high levels of SRSF2 protein (Figure 3.4). Importantly, this result also demonstrates that SRSF2 knock down does not significantly compromise cervical epithelial cell growth. The black line with black squares represents the total cell numbers after the specified time of transfection with a non-E6/E7-targeting fluorescent siRNA

A



B



C

Hours post transfection	W12G		W12GPXY	
	Control	siSRSF2	Control	siSRSF2
24	5.87×10^4	5.77×10^4	6.47×10^4	7.5×10^4
72	1.25×10^5	1.1×10^5	3.58×10^5	1.32×10^5

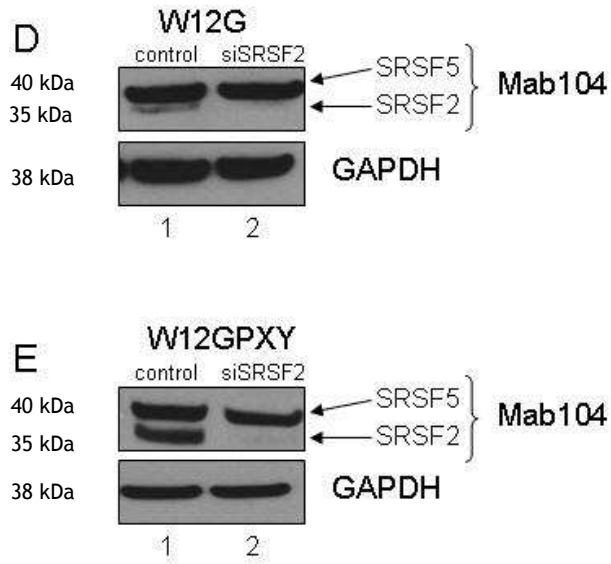


Figure 4.7 SRSF2 knockdown results in decreased cell growth in W12GPXY cells. A) Growth curve analysis of W12G cells treated with siSRSF2 and the total cell numbers counted every 24 hours post transfection. The siSRSF2 cells are shown by the gray line and the control cells indicated by the black line. Graph shows the mean and standard deviation from the mean number of cells. **B)** Growth curve analysis of W12GPXY cells treated with siSRSF2 and the total cell numbers counted every 24 hours post transfection. The siSRSF2 cells are shown by the gray line and the control cells indicated by the black line. Graph shows the mean and standard deviation from the mean number of cells. **C)** Table displaying the actual cell numbers at 24 hours and 72 hours for each treatment. **D)** Western blot showing the knockdown efficiency in W12G cells after SRSF2 knockdown. Mab104 is used to detect SRSF2 levels and Mab6CS is used to detect GAPDH levels. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control. **E)** Western blot showing the knockdown efficiency in W12GPXY cells after SRSF2 knockdown. Mab104 is used to detect SRSF2 levels and Mab6CS is used to detect GAPDH levels. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control.

(Dharmacon). The gray line with gray triangles represents the total cell numbers after the specific times of transfection with SRSF2 siRNA. In contrast to the results with W12G cells, there was a dramatic effect on cell growth rate in the W12GPXY cells upon SRSF2 knockdown (Figure 4.7 B). After 72 hours there was an almost 50% reduction in total cell numbers in the SRSF2 knockdown cells (gray line) compared to the non-targeting fluorescent siRNA treated cells (black line). The experiments were not carried out for longer time periods due to a reduction in effectiveness of the siRNA after this period of time in the transient transfection experiment. 72 hours is the longest recommended time point for transfection, after which the siRNA effect may be lost as the siRNA does not divide with the cells. Moreover, Lipofectamine can be relatively toxic to the cells. This means that any cells remaining after 72 hours may in fact not be transfected and may outgrow the transfected cells. The finding of a slowing of cell growth in W12GPXY cells contributes to the evidence that E6/E7 RNAs and proteins are reduced upon SRSF2 knockdown. These data suggest that SRSF2 is required to maintain growth rate of W12GPXY cells.

When SRSF2 was knocked down in W12GPXY I noticed an alteration in the appearance of the cells. In order to investigate this further, W12GPXY cells were transfected using commercially available SRSF2 siRNA (Dharmacon) and a fluorescent non-targeting siRNA (Dharmacon). Pictures were taken using a Leica light microscope and the 10X objective. Figure 4.8 shows an example of the phenotype of the cells after transfection. Normal untransfected GPXY cells are elongated and grow both in small colonies and in isolation. Lipofectamine RNAiMAX treatment with a non-targeting fluorescent siRNA altered the

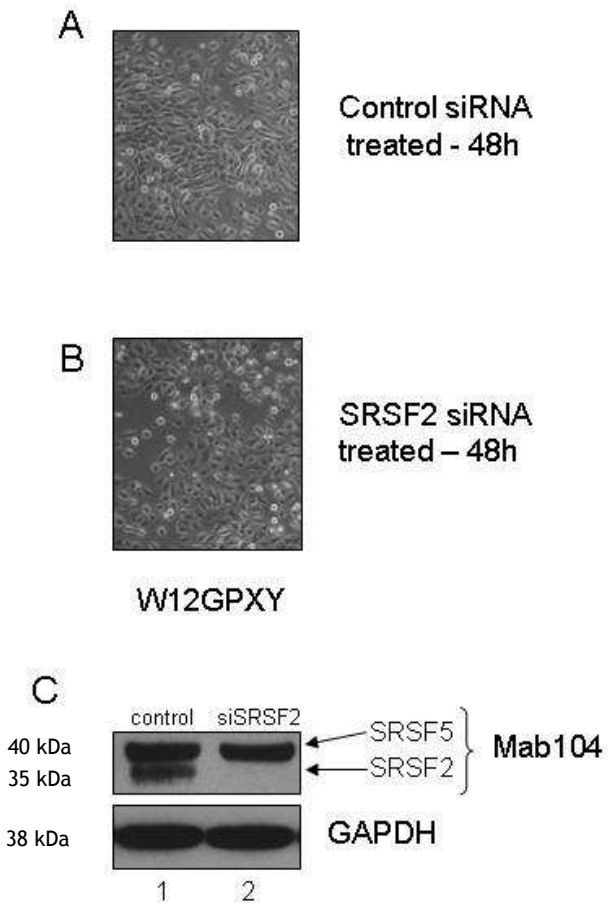
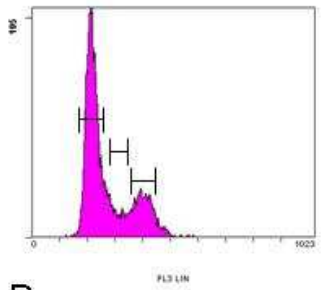


Figure 4.8 SRSF2 knockdown alters the morphology of the transfected cells. **A)** W12GPXY cells transfected with siGlo for 48 hours. Picture was taken with the 10X objective **B)** W12GPXY cells transfected with siSRSF2 for 48 hours. Picture was taken with the 10X objective **C)** Western blot showing knockdown efficiency of SRSF2 compared to control cells. Mab104 is used to detect SRSF2 protein and Mab6CS is used to detect GAPDH protein levels. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control.

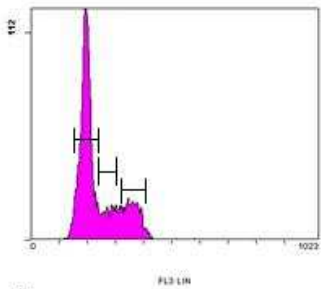
phenotype to some extent with some cells showing a more rounded phenotype. However the cells were still able to grow in large colonies (Figure 4.8 A). When SRSF2 was knocked down however, the cells became more rounded and more individual cells were present (Figure 4.8 B). The pictures were taken 48 hours after transfection. Some of the transfected cells also appeared smaller in size and more refractile than the control-transfected cells. These data suggested that the cells treated with siRNA against SRSF2 were entering senescence or beginning apoptosis.

The observation that W12GPXY cells grow more slowly and alter their cellular phenotype in the presence of reduced SRSF2 levels prompted me to examine cell growth in terms of the cell cycle. As SRSF2 may have oncogenic properties it was possible that cells were growth-arrested in its absence. To examine this I used propidium iodide to stain cellular DNA and examined DNA content in cell populations either untreated, treated with control siRNA or treated with siRNA against SRSF2. W12GPXY cells were mock transfected, transfected with siGLO or with SRSF2 siRNA (Dharmacon) and harvested after 48 hours of transfection. The cells were then fixed in 70% ethanol for 30 minutes on ice before being stained with propidium iodide overnight at 4°C to visualise the DNA content. RNase A was added to digest RNA that could interfere with the DNA profiles. The DNA content was analysed using the EXPO32ADCXL4 Colour program on a Beckman Coulter Epics XL-MCL machine. Figure 4.9 shows the results of one experiment. Two experiments were carried out with very similar results in each case. All three W12GPXY DNA profiles looked similar and each population of cells appear to be cycling normally because distinct G1, S and G2-phase peaks could be observed. There was little difference between the percentage of cells in G1 and S-phase of the cell cycle in the cells mock transfected (Figure 4.9 A) or

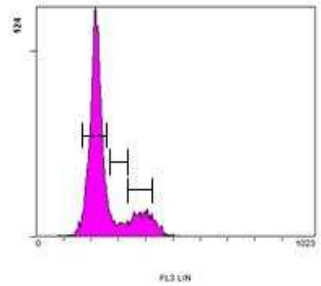
A GPXY mock



B GPXY control



C GPXY siSRSF2



D

Stage in Cell Cycle	mock	control	siSRSF2
Average G_1	65.49%	66.12%	69.21%
Average S	10.4%	12.88%	12.31%
Average G_2	19.87%	15.41%	12.83%

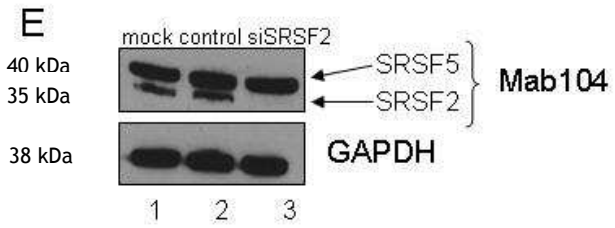


Figure 4.9 SRSF2 knockdown affects the cell cycle. A) DNA content of W12GPXY cells that were not transfected. **B)** DNA content of W12GPXY cells transfected with siGlo for 48 hours. **C)** DNA content of W12GPXY cell transfected with siSRSF2 for 48 hours. **D)** Table displaying the mean number of cells in each stage of the cell cycle in two independent experiments. **E)** Western blot showing knockdown efficiency of SRSF2 compared to control cells. Mab104 is used to detect SRSF2 protein and Mab6CS is used to detect GAPDH protein levels. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control.

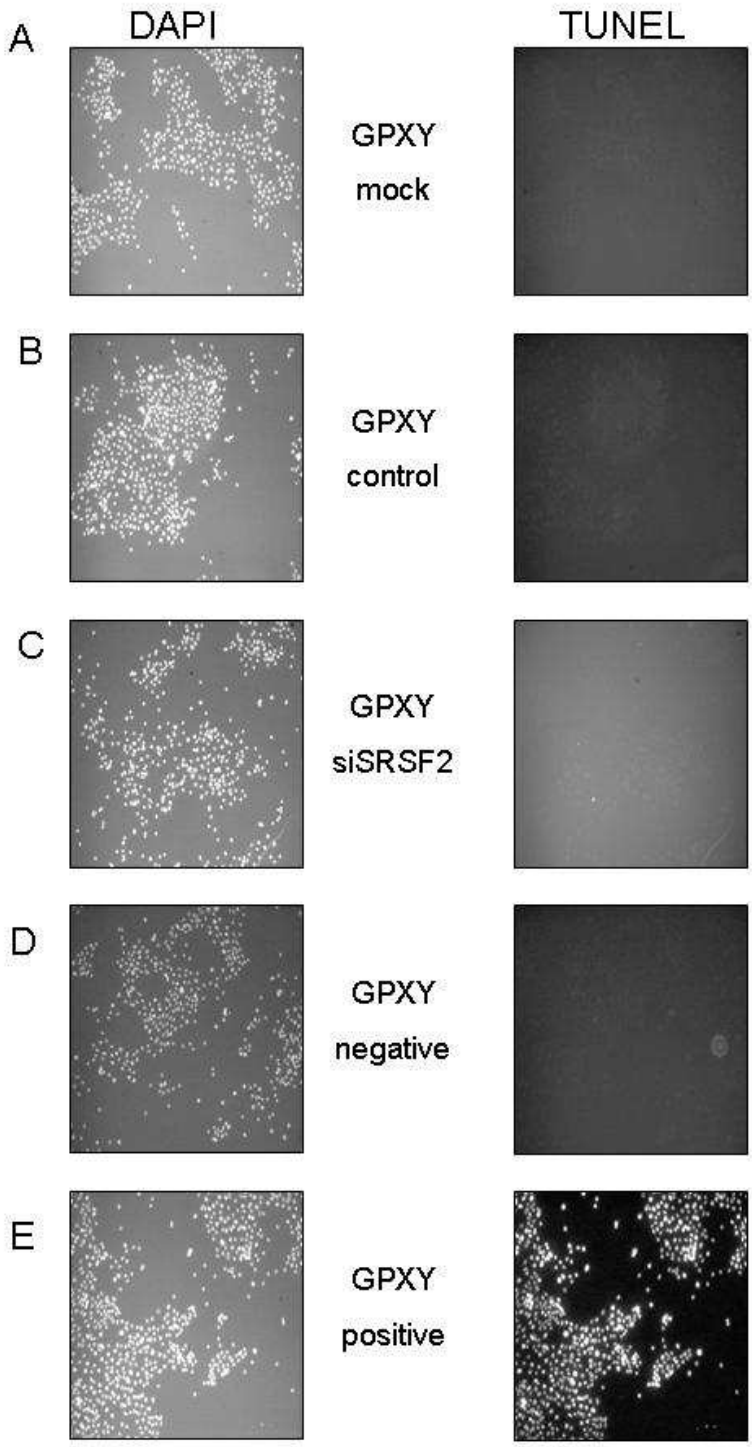
transfected with a fluorescent non-targeting siRNA (Figure 4.9 B) except that a reduction in G2 phase was noted in the latter. However, after SRSF2 knockdown there was around a 7% decrease in the percentage of cells in the G_2 stage of the cell cycle with a corresponding increase of cells in the G_1 stage compared to mock transfected cells (Figure 4.9 C). Figure 4.9 D summarises the quantification of the mean percentage of cells in each stage of the cell cycle in two independent experiments. These data suggest that SRSF2 knock down might arrest the cells in G_1 or at least could slow the cycling time of the cells.

4.1.4.2 SRSF2 knockdown results in apoptosis

As SRSF2 knockdown affects the morphology and growth of the cells and may affect the cell cycle, the next feature to investigate was whether SRSF2 knockdown caused the cells to enter senescence or apoptosis. One observation that suggested that this may be the case, was that knockdown of SRSF2 resulted in visibly more cells detaching from the cell culture dish and floating in the growth medium. Cell detachment could suggest cell senescence or cell death. To examine senescence, Senescence-Associated ß-Galactosidase Assays (SA ß-Gal) (New England Biolabs) were carried out. SA- ß gal staining only occurs at pH 6 therefore only cells that are senescent and have a reduced pH will stain positive for ß- galactosidase. W12GPXY cells were treated with siSRSF2 (Dharmacon) or siGlo (Dharmacon) for 24-72 hours and fixed before being stained for expression of SA-ß-gal. However, regardless of the length of time of SRSF2 knockdown no ß-gal staining was observed (data not shown). One problem with this assay, however is that there was no appropriate standard positive control suggested by the kit, therefore lack of staining could indicate there is no

senescence taking place or that the kit was not working efficiently. Tumour cells do not senesce and I did not have access to any primary cell lines. It has previously been suggested that expressing E2 in HPV16 positive transformed cells will induce senescence in the transfected cells however these experiments have been done with overexpression vectors for E6, E7 and E2 (DeFilippis et al., 2003). Although W12GPXY cells express E6 and E7 they are not highly expressed compared to overexpression constructs. The study also used BPV E2 to repress HPV16 E6 an E7 expression and a recent publication by Ottinger et al (2009) has suggested that different PV LCRS and E2 proteins behave differently depending on the cell type and the combination of LCR and E2. However, I did include an HPV16 E2 transfected control in one experiment and I still observed no staining. Therefore it is possible that my cells do not undergo senescence or that the kit was not working. I therefore moved on to studying apoptosis as there were positive controls available for these assays.

To examine apoptosis, TUNEL assays were carried out. TUNEL (Terminal deoxynucleotidyl transferase mediated dUTP nick end labelling) assays detect DNA fragmentation that occurs during apoptosis. When the DNA is fragmented the fluorescently labelled dUTP nucleotide can be incorporate into the DNA during end-labelling and can be detected by fluorescent microscopy. W12GPXY cells were plated onto 4 well chamber slides and either mock transfected or transfected with control siGLO siRNA or with SRSF2 siRNA (Dharmacon) for 24, 48 and 72 hours. After transfection cells were fixed in 4% formaldehyde at 4°C and permeablised with 0.2% Triton X-100. The cells were stained for 1 hour at 37°C in a humidified chamber before the reaction was stopped. As a staining control,


A
DAPI
TUNEL
GPXY
mock
B
GPXY
control
C
GPXY
siSRSF2
D
GPXY
negative
E
GPXY
positive

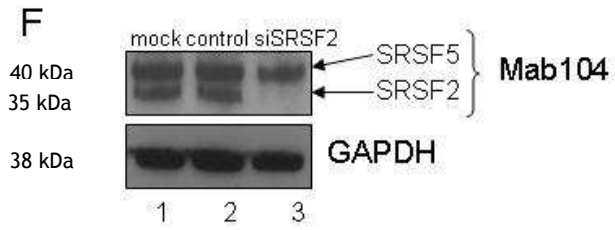
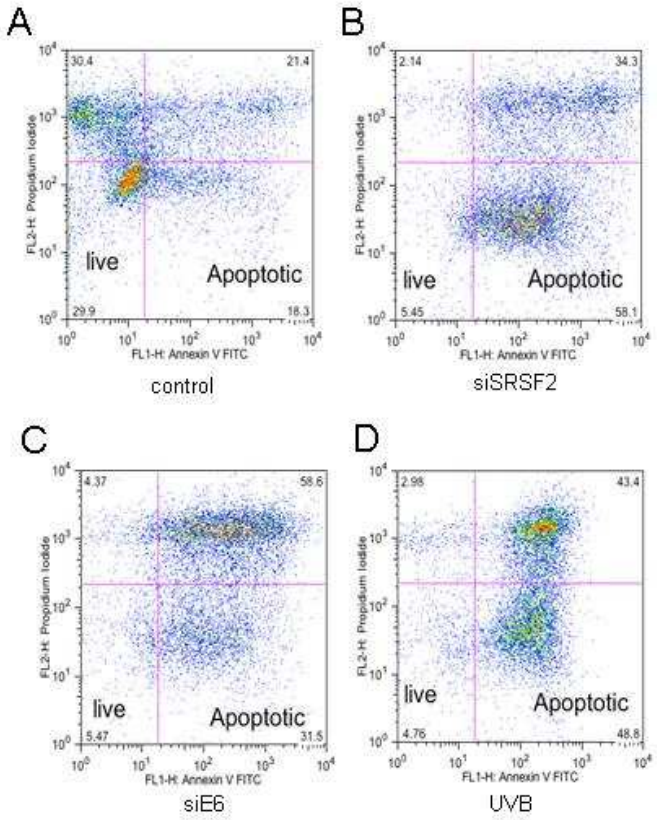


Figure 4.10 Cells with treated with siSRSF2 do not stain positive for TUNEL. **A)** W12GPXY cells without siRNA transfection. **B)** W12GPXY transfected with siGlo for 48 hours. **C)** W12GPXY cells transfected with siSRSF2 for 48 hours. **D)** W12GPXY cells without TUNEL stain added to act as a negative control. **E)** W12GPXY treated with DNase-1 to act as a positive staining control. **F)** Western blot showing knockdown efficiency of SRSF2 compared to control cells. Mab104 is used to detect SRSF2 protein and Mab6CS is used to detect GAPDH protein levels. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control.

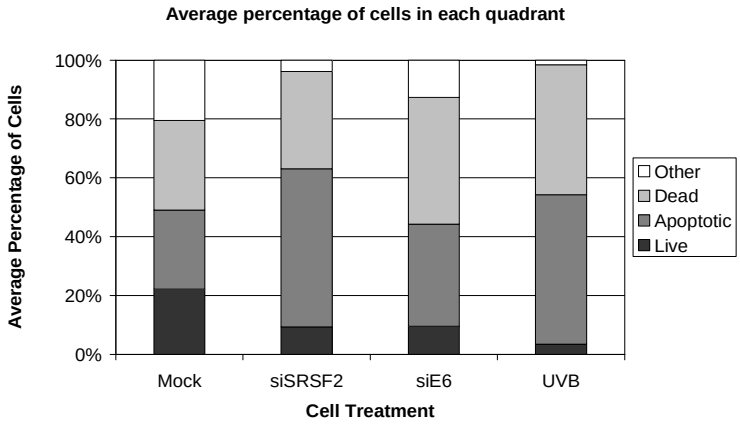
mounting media containing DAPI was used to stain nuclei. A positive control was created by adding DNase-1 to the control cells before staining. A negative staining control with no labelled dUTP added to the staining solution was also included. Figure 4.10 shows an example of the staining achieved after 48 hours of transfection. DAPI staining was detected in all samples showing the number of cells that were present in the assay. Although TUNEL staining could be observed in the positive control W12GPXY cells and no staining was detected in the negative control cells, no TUNEL staining was evident in any of the test W12GPXY cell populations at any time post-transfection. This result indicated that SRSF2 knockdown in W12GPXY cells did not cause apoptosis, which also concurs with the small sub-G1 peak in the Flow Cytometry data in Figure 4.9. However as TUNEL staining detects nicks in the DNA it is only detecting the late stage of apoptosis and will not detect cells that are entering into early apoptosis.

An alternative method of detecting cells undergoing apoptosis uses Annexin V detection. Annexin-FITC conjugate can detect the presence of phosphatidylserine on the surface of cells which is a marker of early apoptosis. Phosphatidylserine is normally present on the inside of the plasma membrane however once the cell enters apoptosis the caspase signalling cascade is activated and the phosphatidylserine is relocated to the external membrane. This allows the apoptotic cell to be identified and cleared in a controlled manner. Therefore Annexin V apoptosis detection often gives a better indication of cells undergoing apoptosis. I designed an apoptosis assay to determine the effect of SRSF2 knockdown that regulates E6/E7 mRNA abundance in W12GPXY cells relative to direct knockdown of E6/E7 RNA. W12GPXY cells were either transfected with control siRNA or transfected with SRSF2 siRNA (Dharmacon) or

with E6 siRNA for 48 hours before being harvested. The E6 siRNA is situated in the first exon and will therefore knockdown all E6 isoforms. The cells were resuspended in Annexin binding buffer and stained with Annexin V 488nm and propidium iodide for 15 minutes at room temperature. A positive apoptosis control was included by treating untransfected W12GPXY cells with 500 J/m^2 of UVB 24 hours prior to harvesting. The cells were then analysed on a BD Biosciences FACSCalibur machine and the results analysed using FlowJo software. Figure 4.11 shows the result of one such experiment. Three experiments were carried out with similar results in each case. For the plots, the lower left quadrant contains cells which are negative for both annexin V and propidium iodide and are therefore live cells. The lower right quadrant contains cells that are positive for annexin V and negative for propidium iodide and are therefore in apoptosis. The upper right quadrant contains cells positive for both annexin V and propidium iodide and are considered to be dead through either apoptosis or necrosis. The upper left quadrant is cells that are positive only for propidium iodide. Anything in this quadrant is most likely cellular debris. Figure 4.11 A shows the cell profile after treatment with a non-targeting fluorescent siRNA. The majority of cells were in the lower left hand side quadrant of the plot representative of live cells. UVB treatment was used as a positive control for cell death and resulted in the majority of the cells entering the upper and lower right hand side of the plot representing dead cells and cells undergoing apoptosis respectively (Figure 4.11 D). Following SRSF2 knockdown the majority of cells entered the lower right hand quadrant indicating that apoptosis was indeed taking place (Figure 4.11 B). Interestingly, the cell population profile after E6 siRNA treatment displayed a similar pattern, although more cells occupied the upper right hand quadrant compared to cells with reduced SRSF2 levels (Figure 4.11 C).


A
B
C
D
live
Apoptotic
FL2-H: Propidium Iodide
FL1-H: Annexin V FITC
control
siSRSF2
siE6
UVB

E


Average percentage of cells in each quadrant
Average Percentage of Cells
100%
80%
60%
40%
20%
0%
Mock
siSRSF2
siE6
UVB
Cell Treatment
Other
Dead
Apoptotic
Live

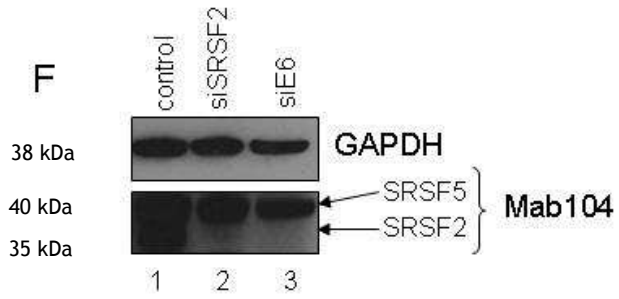


Figure 4.11 Cells with treated with siSRSF2 stain positive for annexin V. A) W12GPXY cells transfected with siGlo for 48 hours and stained with annexin V and propidium iodide. **B)** W12GPXY transfected with siSRSF2 for 48 hours and stained with annexin V and propidium iodide. **C)** W12GPXY cells transfected with siE6 for 48 hours and stained with annexin V and propidium iodide. **D)** W12GPXY cells treated with 500J/cm2 for 24 hours and stained with annexin V and propidium iodide. **E)** Graph displaying the mean percentage of cells in each quadrant after each treatment. Graph shows the results from three independent experiments. **F)** Western blot showing knockdown efficiency of SRSF2 compared to control cells. Mab104 is used to detect SRSF2 protein and Mab6CS is used to detect GAPDH protein levels. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control.

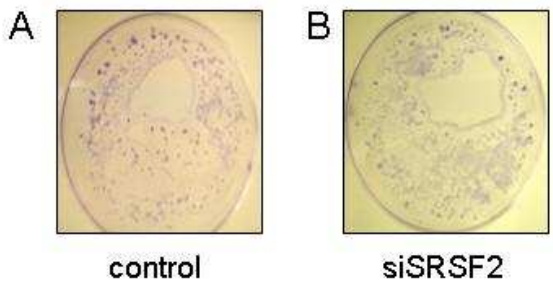
Figure 4.11 E shows the combined results of three experiments. The graph shows the mean percentage of cells in each quadrant from the three independent experiments. There is one error in this experiment. The control siRNA in the transfection control was fluorescently labelled. This may explain why in the control experiment a high percentage of cells appear positive for propidium iodide staining. The siRNA fluorescence is detected in the red channel as is propidium iodide. This experiment suggests that SRSF2 knockdown cells are entering into early apoptosis however as later time points were not used I cannot definitively say that the cells are dying through apoptosis.

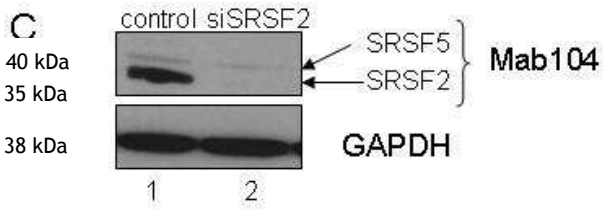
I conclude that knocking down SRSF2 in HPV16 transformed epithelial cells may cause the cells to enter early apoptosis. The annexin V data indicates that the SRSF2 knockdown cells may be entering into early apoptosis, however as the TUNEL data is negative and there is an absence of a sub-G1 peak in the Flow Cytometery analysis late stage apoptosis cannot be confirmed. Therefore it is possible that SRSF2 knockdown cells are becoming positive for Annexin V staining via another route. For example ER stress can result in the release of calcium followed by the plasma memebrane becoming permeabilised and could result in the redistribution of phosphatidylserine and Propidium iodide could then enter to stain the DNA. Apoptosis was suggested due to the altered morphology with the cells rounding up and detaching from the culture dish. However in order to definitively confirm apoptosis further experiments including extending the time points studied in the Annexin V experiment and determination of cell viability by methods such as an MTT assay or Hoescht 33342 staining.

4.1.4.3 SRF2 knockdown inhibits anchorage independent growth

I have shown that knockdown of SRSF2 slows cell growth rate, changes the phenotype of the cells and causes the cells to enter early apoptosis. These observations indicate that SRSF2 is required for maintenance of the tumour phenotype of W12GPXY cells. One final phenotype commonly attributed to tumour cells is the ability to form colonies in soft agar. Knockdown of SRSF3 in A2780 ovarian cancer cells has already been shown to result in a reduced ability to form colonies in soft agar (He et al., 2011) . To discover whether SRSF2 knockdown also resulted in a reversal of anchorage-independent growth W12GPXY cells were transfected with siRNA designed against SRSF2 (Dharmacon) for 24 hours before being harvested and plated out on soft agar. A 24 hour transfection was used as I did not want to loose any of the knockdown cells that may have detached from the cell culture dish when harvesting the cells to plate out onto soft agar. The base layer contained 0.75% agar and the top layer contained 0.4% agar. This provided a concentration gradient of agar. The base layer was added and allowed to set before the cells were added to the top layer and the cell-containing top layer plated out onto the base layer. The cells were allowed to grow for 14 days before the medium was removed and the colonies stained with 0.005% crystal violet for 1 hour before being dried and photographed. Figure 4.12 A shows an example of the appearance of colonies of W12GPXY cells transfected with control siRNA present after staining. A number of colonies can be clearly seen, especially on the periphery of the plate. Figure 4.12 B shows a plate of colonies following transfection with SRSF2 siRNA (Dharmacon). The reduction in SRSF2 levels resulted in a reduced number of colonies: fewer colonies were stained and the size of the colonies that were present was smaller, on average, to that of the control siRNA-transfected cells.

Figure 4.12 D shows a quantification of three independent experiments. Colonies larger than a defined square were counted. The pictures were imported into Powerpoint and adjusted to equal sizes. I zoomed in to 300% and drew the smallest square possible from the Powerpoint tool. The same square was used to count the colonies in all pictures. The experiment was carried out three times and the graph shows the mean number of colonies larger than the square. I conclude from this experiment that SRSF2 knockdown can reverse the transformed phenotype of W12GPXY cells.


A
B
control
siSRSF2


C
control siSRSF2
40 kDa
35 kDa
38 kDa
SRSF5
SRSF2
Mab104
GAPDH
1
2

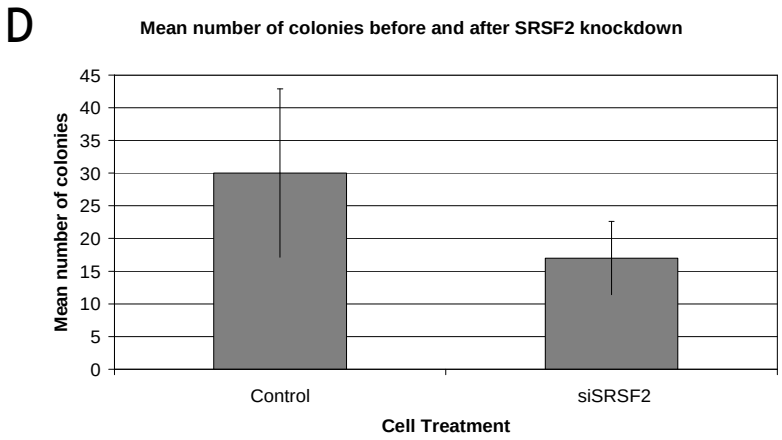
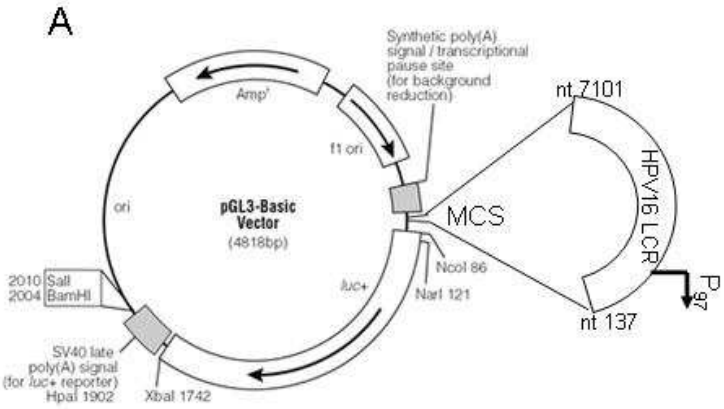

D
Mean number of colonies before and after SRSF2 knockdown
Mean number of colonies
45
40
35
30
25
20
15
10
5
0
Control
siSRSF2
Cell Treatment

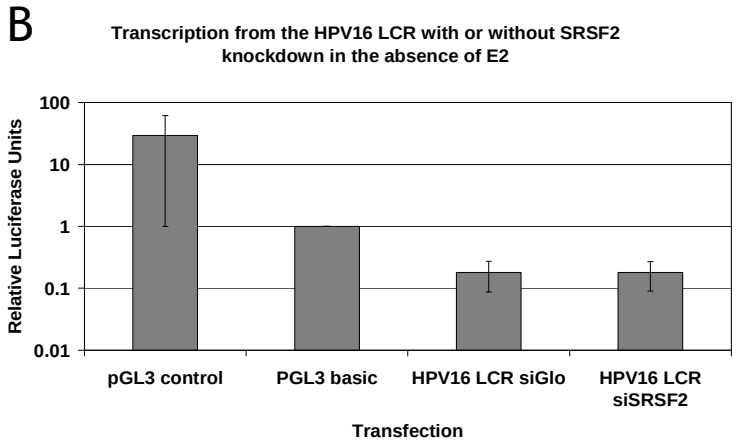
Figure 4.12 Cells with treated with siSRSF2 have a reduced ability to form colonies in soft agar. A) W12GPXY cells transfected with siGlo for 48 hours grown in soft agar for 14 days and stained with 0.005% crystal violet. **B)** W12GPXY transfected with siSRSF2 for 48 hours and grown in soft agar for 14 days and stained with 0.005% crystal violet. **C)** Western blot showing knockdown efficiency of SRSF2 compared to control cells. Mab104 is used to detect SRSF2 protein and Mab6CS is used to detect GAPDH protein levels. The protein being detected is indicated to the right of the blot. GAPDH is used as a loading control. **D)** Graph displaying the mean number of colonies and the standard deviation from the mean of colonies larger than a defined square after each treatment. Graph shows the results from three independent experiments.

Another important feature of tumour cells is their ability to invade a collagen matrix. An invasive phenotype is indicative of a malignantly transformed cell. Therefore to determine if SRSF2 knockdown would also reverse the invasive phenotype of the W12GPXY cells, I performed Matrigel invasion assays. Before starting with transfections, W12GPXY cells were assayed for their ability to invade without any treatment. Matrigel (BD biosciences) was diluted 1:1.5 in serum free DMEM and added to the appropriate number of transwells before being left to set. $2x10^4$ cells were added to the bottom of the transwell and incubated at 37°C for 3 hours to allow the cells to attach. The transwells were washed twice in serum free medium before being placed into 600µl serum free medium. 150µl of DMEM containing 10% FBS was added to the centre of the transwell creating a concentration gradient of medium that will attract any invasive cells to move up through the matrigel towards the serum containing medium. The cells were incubated at 37°C for 3 days before the cells were stained with 4µM Calcein-A for 1 hour at 37°C. Cells were analysed using the Leica Confocal microscope. Unfortunately, W12GPXY cells did not invade into the matrigel matrix even after further dilution of the matrigel matrix and increasing the cell attachment time and cell invasion time. Equine sarcoid control cells that were capable of invading the matrigel gel were used to ensure the protocol was carried out correctly, however W12GPXY cells were unable to invade the matrigel matrix. W12GPXY cells alone are not invasive in a matrigel system therefore it would have been impossible to determine whether knocking down SRSF2 would have any effect on invasivness of the cells. Perhaps another method of determining migration would have been to carry out a similar experiment using a different matrix for example a collagen matrix.

4.1.5 The mechanism of action of SRSF2 on E6/E7 expression: SRSF2 does not trans-activate the P_{97} promoter.

The reduction in E6/E7 RNA after SRSF2 knockdown could be accounted for by a number of different mechanisms. As SRSF2 is thought not to traffic between the nucleus and cytoplasm, it is unlikely to be as a result of changes to SRSF2-mediated trafficking of the RNAs. Moreover, as evidence suggests that SRSF2 is confined to the nucleus it is unlikely to control E6/E7 expression through a cytoplasmic event, for example translational control. Therefore I reasoned that either SRSF2 might exert a transcriptional effect on the P_{97} promoter from which E6/E7 mRNA is transcribed or SRSF2 might alter nuclear stability of the E6/E7 RNA. SRSF2 has previously been shown to promote transcriptional elongation (Lin et al., 2008). SRSF2 was found to be associated with the Pol II C-terminal domain (CTD) subunit and P-TEFb which phosphorylates serine 2 of Pol II CTD to allow for transcription elongation. When SRSF2 was depleted the phosphorylation at serine 2 was reduced and the total mRNA levels were reduced (Lin et al., 2008). SRSF2 has also recently been shown to stabilise a target mRNA containing a SRSF2 binding site. SRSF2 was shown to bind to tau mRNAs containing exon 10 and stabilise the RNAs (Qian et al., 2011). The authors demonstrated that the increase in mRNAs was not due to increased transcription but by direct binding and stabilisation of the mRNAs (Qian et al., 2011). Either of these scenarios could be occurring with regards to the E6/E7 mRNAs.


A
Synthetic poly(A) signal / transcriptional pause site (for background reduction)
Amp
f1 ori
ori
pGL3-Basic Vector (4818bp)
MCS
nt 7101
HPV16 LCR
P97
nt 137
Ncol 86
Narl 121
luc+
2010 Sall
2004 BamHI
SV40 late poly(A) signal (for luc+ reporter)
HpaI 1902
XbaI 1742


B
Transcription from the HPV16 LCR with or without SRSF2 knockdown in the absence of E2
Relative Luciferase Units
100
10
1
0.1
0.01
pGL3 control
PGL3 basic
HPV16 LCR siGlo
HPV16 LCR siSRSF2
Transfection

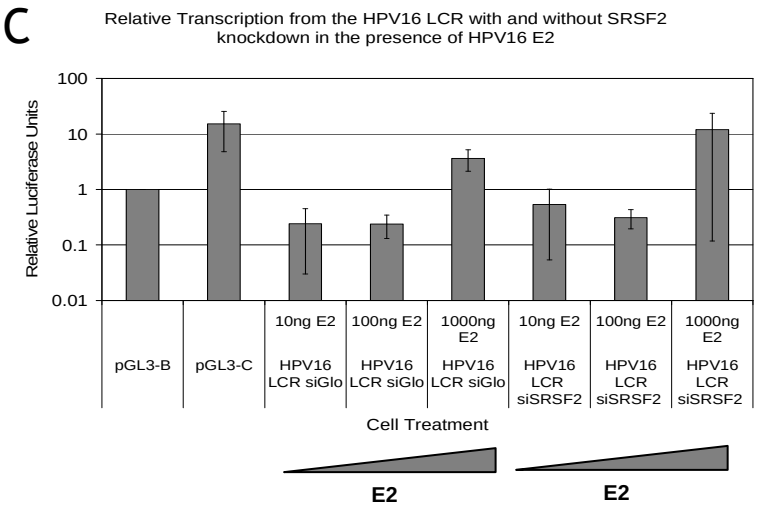

C
Relative Transcription from the HPV16 LCR with and without SRSF2 knockdown in the presence of HPV16 E2
Relative Luciferase Units
100
10
1
0.1
0.01
pGL3-B
pGL3-C
10ng E2 HPV16 LCR siGlo
100ng E2 HPV16 LCR siGlo
1000ng E2 HPV16 LCR siGlo
10ng E2 HPV16 LCR siSRSF2
100ng E2 HPV16 LCR siSRSF2
1000ng E2 HPV16 LCR siSRSF2
Cell Treatment
E2
E2

Figure 4.13 SRSF2 knockdown does not reduce transcription from the P97 promoter. A) Schematic diagram of the HPV16 LCR construct cloned into pGL3-Basic. The start and end nucleotide of the LCR fragment is indicated as is the approximate position of the major early promoter P97. **B)** Graph showing the results from three independent experiments where 293-T cells were transfected with siGlo or siSRSF2 and HPV16 LCR vector in the absence of E2 for 48 hours. Cells were harvested and the expression of the luciferase reporter measured. The mean and standard deviation from the mean are plotted on the graph. **C)** Graph showing the results from two independent experiments where 293-T cells were transfected with siGlo or siSRSF2 and HPV16 LCR vector in the presence of E2 for 48 hours. Cells were harvested and the expression of the luciferase reporter measured. The mean and standard deviation from the mean are plotted on the graph.

To test the first hypothesis, an HPV16 LCR transcription assay construct (a gift from Prof Iain Morgan) was transfected into HEK 293-T cells. The construct contained the HPV16 LCR from nucleotide 7101 to +137 and was cloned into pGL3-basic (Promega). The HPV16LCR construct was transfected into 293-T cells with increasing concentrations of an HPV16 E2 expressing construct (a gift from Prof Peter Howley) with or without SRSF2 knockdown. A schematic diagram of the LCR construct is shown in Figure 4.13 A. The HPV16LCR construct contains a luciferase reporter plasmid, therefore transcription from the P_{97} promoter can be quantified by levels of luciferase produced after transfection. The cells were transiently transfected using Lipofectamine (Invitrogen) and incubated for 48 hours before harvesting. The transfected cells were lysed in Reporter lysis buffer (Promega) and the levels of luciferase determined by spectrophotometer. pGL3-B is a control plasmid which does not have a promoter therefore it is used as a background luciferase control and any expression above the pGL3-B levels is an activation of luciferase transcription. pGL3-C is another control plasmid which acts as a positive control. In this plasmid luciferase is under the control of a strong CMV promoter therefore luciferase should be highly expressed, this provides a control for the detection of the luciferase by the spectrophotometer. In Figure 4.13 B, I first carried out transcription assays using only the HPV16 LCR construct without the addition of E2 as I thought perhaps any action by E2 would mask any effects on the LCR activity after SRSF2 knockdown. From this experiment there is no significant difference between the siGlo and siSRSF2 treated samples; the luciferase production is almost equal. However it is clear from this experiment that the HPV16 LCR is not active in these cells as the levels of luciferase expression is below that of the pGL3-B negative control. I then went on to carry out the same experiment with the addition of E2. A titration of E2 concentration was used as E2 has been reported to have activation and

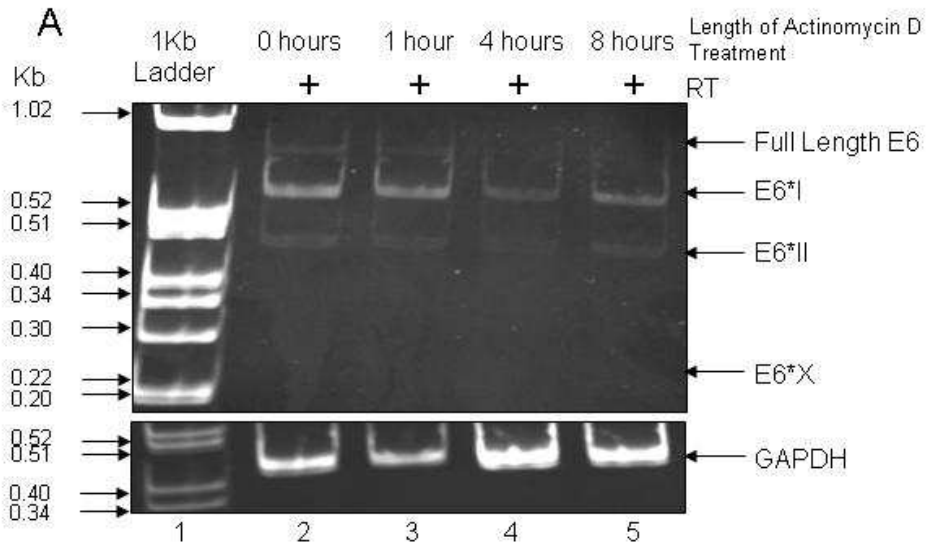
repression activities depending on the concentration. Therefore a titration was used to ensure optimal activation of the HPV16 LCR. Again pGL3-B and pGL3-C are used as controls and shown in Figure 4.13 C. From this experiment it does not seem like SRSF2 knockdown is having any effect on LCR activity. When 10ng of E2 is added to the cells the levels of transcription in both the siSRSF2 and siGlo treated cells are very low and there is no significant difference between the luciferase expression levels. Similarly when 100ng of E2 is added the levels of transcription are again very low with no difference between the siGlo and siSRSF2 samples. This suggests that at these concentrations of E2, E2 does not activate transcription of the HPV16 LCR and may in fact repress the LCR as the levels of transcription are below that of the negative control pGL3-B construct. However, when 1000ng of E2 is added there is a little activation from the HPV16 LCR. Although there is activation there is once again no significant difference between the siSRSF2 and siGlo knockdown cells. The results from the transcription assay are difficult to draw any conclusions from. Knockdown of SRSF2 does not seem to have a detrimental effect on transcription factors as the levels of transcription after SRSF2 knockdown are similar to those without knockdown. If anything transcription could be being activated although this is far from convincing. Therefore unfortunately it is not possible to say what effect SRSF2 has on transcriptional control from these experiments. These results are similar to a recent publication by Ottinger et al (2009). They reported that HPV16 LCR activity is very weak in human cells when compared to other high and low risk papillomaviruses (Ottinger et al., 2009). The authors also reported that LCR activity for all of the different LCRs investigated were slightly different depending on the cell type used (Ottinger et al., 2009). Perhaps 293-T cells do not provide the HPV16 LCR with the required environment for activity. A previous study also reported that high levels of E2 could repress

the HPV16 LCR only when the HPV genome is integrated and not in the episomal form (Bechtold et al., 2003). This may be why I do not see repression of the HPV16 LCR with very high concentrations of E2.

4.1.6 The mechanism of action of SRSF2 on E6/E7 expression: SRSF2 stabilises E6/E7 RNAs

The alternative hypothesis that I addressed is that SRSF2 is required for stability of the E6/E7 RNAs. To test this hypothesis, W12GPXY cells were transfected with control siRNA or with siRNA (Dharmacon) against SRSF2 for 24 hours. Following this 10μg/ml Actinomycin D was added to inhibit *de novo* RNA synthesis. A time course of Actinomycin D treatment was carried out to determine whether the stability of the E6/E7 RNA was altered upon SRSF2 knockdown. RNA was harvested at 0, 1, 2, 4 and 8 hours following addition of actinomycin D and DNase-1 treated. RNAs were reverse transcribed and amplified using E6/E7 primers described in section 3.1.1. Semi-quantitative PCRs were carried out and GAPDH amplification was used as an internal control standard. The PCR products were separated on a 6% acrylamide gel. The gel was stained with ethidium bromide to visualise the DNA. The band intensity of the E6*I isoform was measured and the results of two independent experiments shown in the graph (Figure 4.14). E6*I was measured as after the later time points this was the only visible isoform as it is the most highly expressed isoform. There was little change in E6 RNA levels between 1-8 hours of actinomycin D treatment indicating that E6/E7 mRNA is quite stable in the W12GPXY cells with little fall-off in levels over the 8 hours time course. This finding corresponds to a report

that integration of HPV16 E6/E7 gene region into host chromosomes results in a longer mRNA half life (Jeon & Lambert, 1995). The lower starting level is probably due to an error in loading the acrylamide gel. After knockdown of SRSF2 a clear difference was observed. After 1 hour incubation in the presence of actinomycin D the levels of E6/E7 RNA decrease significantly and remain low until the end of the time course period at 8 hours indicating that the half life of the E6*I RNA is significantly reduced. These data suggest that SRSF2 controls the stability of the E6/E7 transcripts and is required for high levels of oncoprotein RNA.


A
1Kb Ladder
0 hours
1 hour
4 hours
8 hours
Length of Actinomycin D Treatment
Kb
RT
1.02
0.52
0.51
0.40
0.34
0.30
0.22
0.20
Full Length E6
E6*I
E6*II
E6*X
GAPDH
1
2
3
4
5

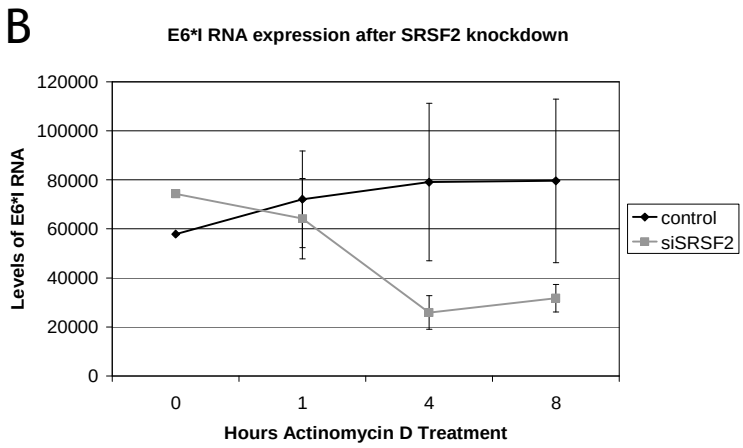

B
E6*I RNA expression after SRSF2 knockdown
Levels of E6*I RNA
120000
100000
80000
60000
40000
20000
0
0
1
4
8
Hours Actinomycin D Treatment
control
siSRSF2

Figure 4.14 SRSF2 knockdown alters the stability of the E6 isoform RNAs. A) Upper panel Ethidium-bromide stained acrylamide gel electrophoresis of E6/E7 isoforms RT-PCR amplified from Actinomycin D treated siSRSF2 knockdown RNA from W12GPXY cells. Isoforms are indicated to the right of the gel. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "O hours", no Actinomycin D added RNA; "1 hour", RNA treated for 1 hour with Actinomycin D; "4h", RNA treated for 4 hours with Actinomycin D; "8h", RNA treated for 8 hours with Actinomycin D; Lower Panel Ethidium-bromide stained acrylamide gel electrophoresis of GAPDH RNA RT-PCR amplified from Actinomycin D treated siSRSF2 knockdown RNA from W12GPXY cells. RT, reverse transcriptase; "-", RT-PCR reaction in the absence of reverse transcriptase; "+", RT-PCR reaction in the presence of reverse transcriptase; "O hours", no Actinomycin D added RNA; "1 hour", RNA treated for 1 hour with Actinomycin D; "4h", RNA treated for 4 hours with Actinomycin D; "8h", RNA treated for 8 hours with Actinomycin D. **B)** Graph showing the results from two independent experiments where W12GPXY cells were transfected with siGlo or siSRSF2 and 10μg/ml actinomycin D was added for varying lengths of time. E6*I RNA levels were measured by pixel density and plotted on the graph. The gray line represents the W12GPXY cells treated with siSRS2 and the black line represents W12GPXY cells transfected with siGlo. The mean and standard deviation from the mean are plotted on the graph.

4.2 Discussion

Results from Chapter 3 allowed me to formulated the hypothesis that the higher levels of the small isoforms E6*II and E6*X in transformed cells and their apparent growth-promoting activity indicated that they may be transformation associated isoforms. Moreover, overexpression of SR proteins in cervical tumour cells could potentially lead to changes in E6/E7 isoform splicing pattern. To address these hypotheses, individual SR proteins were knocked down and the splicing pattern of E6/E7 analysed. However, knocking down the overexpressed SR proteins did not result in an obvious change in E6/E7 isoform production. This was surprising because bioinformatic analysis of the E6/E7 coding region indicated possible high affinity binding sites for these SR proteins that might indicate control of alternative splicing of E6/E7 mRNAs by these proteins due to their close proximity to splice junctions (Figure 4.1). The major finding from these experiments was that knocking down SRSF2 resulted in a reduction of total levels of E6 RNAs (Figure 4.2 D). Importantly, this reduction in E6/E7 RNAs was not restricted to W12GPXY cells, which were derived in the laboratory, but also occurred in CaSki cells which were isolated directly from a cervical cancer patient. Significantly, the reduction in E6/E7 RNA in both these cell lines could indicate that the SRSF2 control over E6/E7 RNA expression may be common in tumours initiated by HPV16 E6/E7 overexpression.

My results suggest that SRSF2 is involved in processing of the E6 RNAs at some stage in the nuclear RNA biogenesis pathway because loss of the protein caused destabilisation of the E6/E7 RNAs. SR proteins are involved in almost every

stage of the RNA maturation process and interference with any stage may induce nuclear degradation of RNA. Therefore the mechanism of action of SFSF2 could be at any number of stages. However, it is clear that the destabilised E6/E7 RNAs are spliced and polyadenylated as several RNA isoforms are amplified after reverse transcription using an Oligo dT primer, so one possibility is that SRSF2 is acting to protect mRNAs from nonsense-mediated decay. The E6/E7 bicistronic mRNA is unusual in that it does not possess a 5' untranslated region and there are several splice acceptor sites close to stop codons (Figure 4.15). How substrates are targeted to the nonsense mediated decay pathway remains somewhat controversial. The original hypothesis was that it was the exon-junction complex positioned at an exon-exon junction close to a termination codon that triggered decay of the RNA, however some recent reports have suggested that it is the distance between the termination codon and the polyadenylation site that triggers decay (reviewed by Brogna & Wen., 2009). The sometimes contrasting findings surrounding nonsense-mediated decay research has led Brogna and Wen to offer their own model which suggests that that it may be early release of the ribosome at a premature termination codon that results in reduced mRNA stability. Under normal circumstances the mRNA would be translated in a closed loop structure and the ribosome would complete several rounds of translation. A premature termination codon would prevent this. This model was proposed to explain why some studies found that the distance between the stop codon and the poly A tail to be the determiner for NMD and others found the presence of an intron downstream from the stop codon to be important for directing NMD (Brogna & Wen, 2009). As during transformation the HPV genome is inserted into the host genome E6 and E7 are no longer processed from the viral polyadenylation site, but from a genomic polyadenylation site downstream from the point of insertion. Therefore either

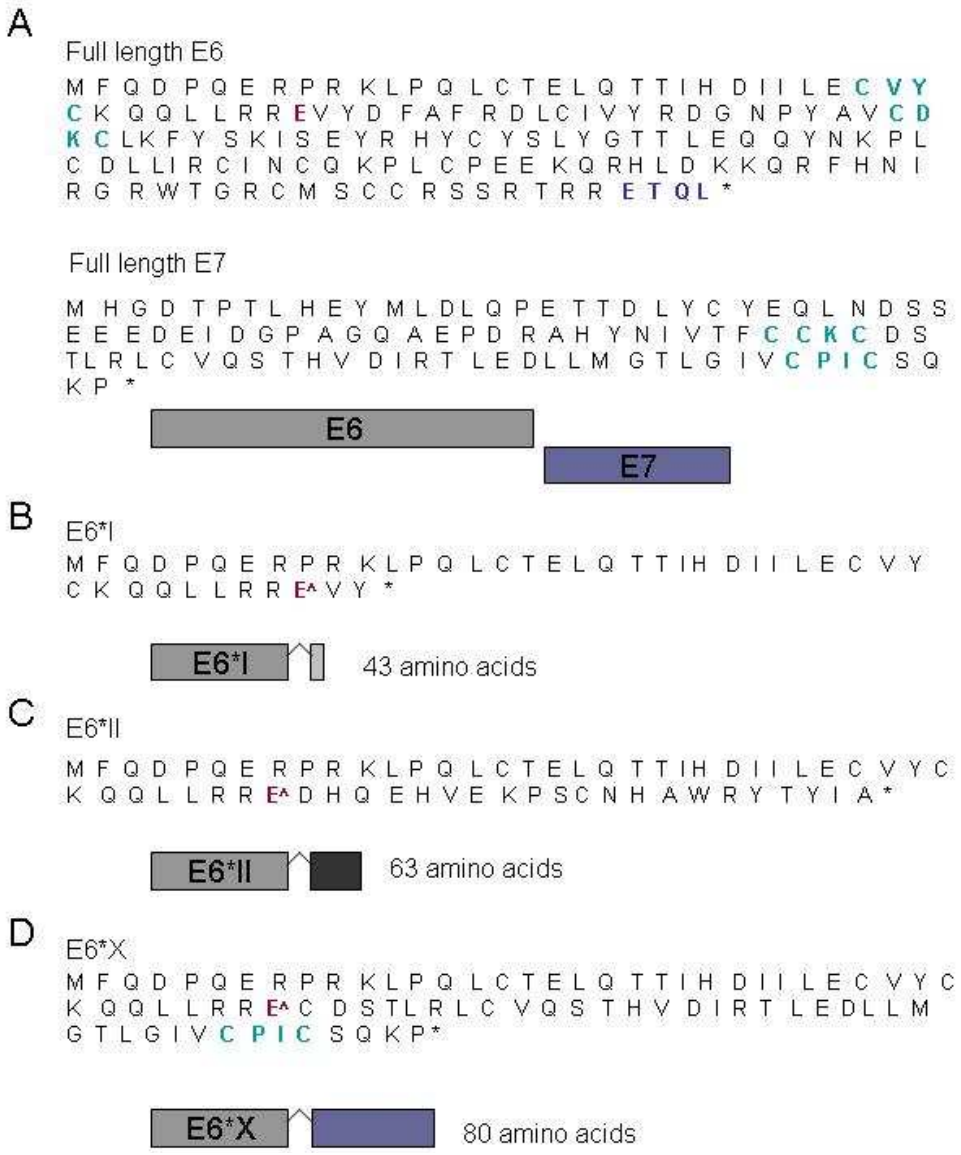


Figure 4.15 Amino acid sequences of the E6 isoforms and small schematic diagrams of the possible proteins. A) Amino acid sequence of full length HPV16 E6 and E7 with the splice donor site highlighted in red. Zinc finger domains are shown in green and the PDZ domain highlighted in blue. **B)** Amino acid sequence of E6*I and diagram of putative protein. **C)** Amino acid sequence of E6*II and diagram of putative protein. **D)** Amino acid sequence of E6*X and diagram of putative protein.

of these NMD mechanisms could be true for E6/E7 RNAs. The E6/E7 open reading frame may be inserted into an open reading frame containing downstream introns or be inserted a long distance from the first genomic polyadenylation site. SRSF2 could perhaps then be recruiting factors to the E6/E7 mRNAs in the nucleus that protect them from NMD in the cytoplasm due to the presence of premature stop codons.

SRSF2 does not however, have a similar effect on E6/E7 mRNA abundance in W12E cells that support the virus life cycle as it does in tumour cells where the life cycle is aborted. When SRSF2 is knocked down in W12E cells, where the HPV16 genome is episomal and virus particles can be produced, there is only a small effect on E6/E7 RNA levels. However when SRSF3 is knocked down in these cells there is a reduction in E6/E7 RNA levels similar to the SRSF2 effect in transformed cells. This is likely due to the polycistronic nature of the viral RNAs during the life cycle. In transformed cells the only viral RNAs normally expressed are E6, E7 with little or no other viral RNA being produced, whereas in W12E cells E6 and E7 are expressed as part of RNAs that also contain sequences for some other viral RNAs including the late RNAs L1 and L2 (Milligan et al., 2007). Therefore other RNAs may require other SR proteins for their processing and those SR proteins may have stronger affinities for the viral RNAs.

SRSF3 has already been shown to be required for E6/E7 RNA production in transfected U2OS cells (Jia et al., 2009). However this study was carried out in an osteosarcoma cell line (U2OS) using constructs created from the late HPV16 promoter p670 to the late polyadenylation site under control of a CMV promoter. The use of alternative promoters and polyadenylation sites will alter the

processing of the resulting RNAs and will affect the splicing activity. My experiments were all carried out in cervical epithelial cells with naturally occurring viral genomes which are under control of the viral promoters and polyadenylation sites.

Not only does reducing SRSF2 protein levels in HPV16 transformed cells reduce E6/E7 RNA levels, but E6 protein levels are also reduced as indicated by the increase in p53 levels in Figure 4.5. This is significant because if p53 levels are increased the cells can then activate Bax and caspase 9 to enter apoptosis. p53 inactivation or mutation is a common feature of a number tumours therefore restoring p53 signalling could reverse the transformed phenotype of the cells.

Overexpression of SRSF2 in these cells at least seems to contribute towards the transformed phenotype of the cells. FACS analysis of apoptosis in W12GPXY cells treated with siRNA against SRSF2 showed that these cells were in early apoptosis suggesting that SRSF2 contributes towards the tumourigenic phenotype. In contrast, siRNA knockdown of E6/E7 mRNA caused a high percentage of the cells to fully apoptose. SRSF2 may contribute to the tumourigenic phenotype by maintaining high levels of E6/E7 expression or it may have other growth-promoting, anti-apoptotic activities through alterations it makes in splice isoform production of oncogene or tumour suppressor RNAs. siRNA knockdown studies indicated SRSF2 overexpression was promoting cellular proliferation and anchorage independent growth. All of these features are hallmarks of tumour cells and knockdown of SRSF2 also appeared to affect the cell cycle progression of the transfected cells. Knockdown of SR proteins SRSF1 and SRSF2 have already been shown to arrest the cell cycle in tumour cells. However the

reported arrest occurred at the G_2/M checkpoint (Li et al., 2005, Xiao et al., 2007) whereas my data suggested an accumulation in G1 (Figure 4.9) . It has also recently been shown that SRSF1 and SRSF3 bind to chromatin in interphase, dissociate during mitosis and reattach after completion of mitosis (Loomis et al., 2009). It is thought that these SR proteins could be directing the dissociation of the HP1 from the chromatin and therefore allowing proper condensation and chromosome segregation to occur during mitosis (Loomis et al., 2009). HP1 is part of the Heterochromatin protein family and can bind to chromatin and repress the expression of genes. The finding of a different cell cycle arrest point in my study does not fit the above pattern. However, if depletion of SRSF2 causes a subsequent increase in p53 levels then this could cause arrest of the cell cycle in G_1/S phase (reviewed by (Vogelstein et al., 2000) as well as in G_2.

One cannot distinguish whether it is the overexpression of SRSF2 itself that has the transforming effects on the cell, or whether it is the overexpression of SRSF2, and the resulting increase in E6/E7 RNA stability and the likely increase in E6 and E7 protein that is causing the transformation. One can assume that it may in fact be a combination of the two. SR proteins have already been shown to have the capacity to transform cells by altering cellular gene expression and of course E6 and E7 have well documented transforming abilities. Therefore in my view it is likely that SRSF2 overexpression is altering cellular gene expression and stabilising E6/E7 RNAs and that both of these effects are having an oncogenic effect on the cell.

Another aspect of my study and similar studies by others is that the mechanism by which SR proteins become overexpressed in cancer cells is not understood. From the gene copy number experiments in section 3.1.4 of Chapter 3, I can

conclude that the SRSF2 gene is not amplified in the W12GPXY cervical cancer cells. This means that it is likely that there is transcriptional upregulation of expression of this gene. How this happens is not understood. Perhaps viral proteins, for example E6, are activating the transcription factors that activate the SRSF2 promoter. For example bioinformatic analysis of the SRSF2 promoter has identified that E2F among other transcription factors can bind to the SRSF2 promoter and activate or repress transcription of SRSF2. E7 interaction with pRb releases the negative control over E2F transcriptional activity and could be activating SRSF2 expression. Alternatively, perhaps the insertion of the viral DNA into the host cell genome has disrupted either a negative regulator of SRSF2 expression or the viral DNA has inserted into genomic DNA upstream of the SRSF2 gene and caused an activation of transcription of the gene. The second possibility seems less likely as the SRSF2 effect on E6/E7 RNA levels is seen in two completely independent cell lines where insertion sites are different. It may also be the case that as E6 and E7 initiate cell proliferation and DNA synthesis, SR proteins are upregulated by normal mitogenic signalling cascades because they are required at increased levels to carry out a higher volume of RNA processing. I believe that the last point seems the most likely. During transformation the cervical epithelium becomes filled with basal-like proliferating epithelial cells and differentiation is switched off. Mole et al. (2009) already illustrated this in staining of cervical sections. SRSF1 staining in low grade CIN1 cervical sections displayed high levels of SRSF1 in the basal and spinous layer of the epithelium (Mole et al., 2009a). However in high grade CIN 3 lesions the entire epithelium stained positive for SRSF1 (Mole et al., 2009a).

5 Discussion

5.1 General Perspectives

There is emerging evidence that RNA processing factors can have tumour promoting properties. Previously, two SR proteins, SRSF1 and SRSF3 have been shown to be oncogenic, in breast and ovarian cancers respectively (Karni et al., 2007, He et al., 2011). The work in this thesis comprises the first demonstration of oncogenic activities for a third SR protein, SRSF2 and the first demonstration of SR protein oncogenicity in cervical cancer cells. My data suggest that SRSF2 overexpression in cervical epithelial cells is in fact oncogenic and contributes to the transformed phenotype. Reducing SRSF2 overexpression in cervical epithelial cells reverses the transformed phenotype and leads to a resulting decrease in HPV16 E6/E7 RNAs. HPV16 E6 and E7 have well documented tumour promoting properties and stabilisation of the E6/E7 RNAs by SRSF2 may be responsible for promoting the transformation of the cells. However, it is equally possible that transformation of the epithelial cells may be due to the overexpression of SRSF2 itself and the resulting likely alterations in alterative splicing, such as those previously observed for SRSF1, that are responsible for the tumourigenic properties of the cell. Expression of E6 and E7 may be additive to this effect or vice versa: a combination of both possibilities is likely driving transformation.

I have shown that knockdown of SRSF2 reduces the proliferative capacity of the transformed cervical epithelial cell, inhibits anchorage independent growth and sensitises the cells to apoptosis. All of these features are the hallmark of cancer cells. The identification of a single protein being responsible for promoting the

transformed phenotype is surely beneficial for treatment considerations as there are already known drugs that inhibit SR protein function.

My demonstration that the smaller E6 isoforms, E6*II and E6*X, are capable of promoting cell growth is an important result as these are the isoforms that are increased in expression in the transformed cervical epithelial cells, where SRSF2 is overexpressed. The variation in the ability of the isoforms to promote anchorage independent growth is however perhaps unsurprising. It is likely that there is a combinatorial effect of isoform expression that acts to promote colony formation in soft agar. In fact, E7 is surely partly responsible for this phenotype due to its interaction with p600 and subsequent avoidance of anoikis (Huh et al., 2005). Therefore the probability is that E7-expressing isoforms in combination with other E6 isoforms are required for anchorage independent growth.

My data indicating that HPV11 E2 as well as HPV16 E2 upregulates expression of SRSF3 suggests that SRSF3 may in fact be generally required for completion of the HPV life cycle and is not type-specific. Whether it is the alternative splicing functions or other RNA processing functions of SRSF3 that HPV requires has still to be determined. Interestingly, data from our laboratory suggests that expression of SRSF3 is required for expression of HPV16 L1 (T. Klymenko and SV. Graham unpublished information). This may be why HPV16 E2 promotes the expression of SRSF3. If SRSF3 is required for capsid mRNA processing this could also be why low risk viruses also upregulate SRSF3. The fact that HPV11 E2 promotes higher expression of SRSF3 than HPV6 E2 cannot be explained at present unless HPV6 requires a different SR protein for viral RNA processing. This however seems unlikely as HPV6 and HPV11 are closely related viruses as indicated by Figure 1.1.

My finding that SRSF3, but not SRSF1, is required for E6/E7 RNA processing during viral infection contrasts with the report of Somberg and Schwartz (2010) who showed a requirement for high levels of SRSF1 for E6/E7 mRNA production. However their studies were carried out using a subgenomic construct containing the early region of the HPV16 viral genome under control of a strong CMV promoter transiently transfected into HeLa cells along with a plasmid that allowed expression of SRSF1 lacking an RS domain. When the SRSF1 plasmid was co-transfected with the HPV16 early region construct, E6 RNA levels appeared to drop (Somberg & Schwartz, 2010). There are a number of problems with these experiments. Firstly, the overexpression of two plasmids under strong promoters simultaneously could be saturating the transcription machinery and could explain the decrease in mRNA levels. Secondly, the nature of the promoter used can greatly influence the choice of splice site in the RNA leading to production of alternative isoforms that may be unstable. Thirdly, HeLa cells express an increased level of SRSF1 protein as they are fully transformed cells. Finally as HeLa cells are transformed they do not differentiate, therefore even if SRSF1 is in fact acting on the E6/E7 RNAs expressed from the virus early region-containing construct one cannot assume that this would occur during a natural infection of differentiating epithelial cells. The studies carried out during the course of this PhD have been carried out in naturally infected, untransformed cervical epithelial cells which can undergo epithelial differentiation and therefore allow completion of the full viral life cycle. While overexpression studies can be useful for investigating protein-protein interaction, care must be taken when equating the findings to the natural system.

Similarly, contrary to my results, Jia et al. (2009) reported that high levels of SRSF3 are required for E6/E7 RNA processing in transformed epithelial cells.

Their studies were performed in HeLa and CaSki cells. HeLa cells are HPV18-positive and CaSki cells are HPV16-positive. As both cell lines have been culture-adapted for many decades they have likely accumulated many genetic mutations and changes that make them very different from the original tumours from which they were derived. The presence of HPV E6 and E7 expression will aid in the genetic instability and allow the accumulation of mutations. Our W12GPXY cells have not been in culture for long: those that I used were passage 1-6 since being isolated from a squamous cell carcinoma upon injection of the originally-derived W12GPXY cell line into nude mice (Aasen et al, 2003). It is likely that these cells are more similar to the transformed cells of a cervical lesion as they have not undergone extensive culturing and have not had as long to accumulate mutations. Jia et al. (2009) investigated knockdown of SRSF3 in the HPV16 and 18 transformed cells and found a decrease in E6/E7 RNA levels in northern blotting after SRSF3 knockdown. They did not however analyse isoform production or relate their findings to RNA processing. Their northern blot analysis shows multiple bands using an E6/E7 spanning probe for HPV16 and an E7 probe for HPV18. The HPV16 probe sits from 442-816 meaning only full length E6 and E6*I isoforms will be detected by the probe. The northern blot shows the presence of two bands. While there is a reduction in both E6/E7 RNA bands upon SRSF3 knockdown, the authors estimated this to be around 54% reduction. I estimate my E6/E7 total RNA (all isoforms) reduction to be around 70% after SRSF2 knockdown in W12GPXY cells. In CaSki cells I would estimate the reduction in total E6/E7 RNAs after SRSF2 knockdown to be similar to that observed by Jia et al. (2009). Additionally, the SRSF3 knockdown achieved by Jia et al. (2009) was around 95-100%. Perhaps a complete SRSF3 knockdown is not tolerated well by CaSki cells and RNA processing in general is reduced. In

my experiments I ensured that I examined effects of SRSF2 knockdown at around 50% to reduce any catastrophic effects on global cellular splicing.

Over expression of SR proteins is common to a number of cancers and could perhaps be considered as a possible therapeutic treatment target. While reducing SR protein overexpression has been shown to reverse the tumour phenotype, the SR protein that is upregulated seems to be different depending on the tumour type. Therefore, some patient-specific and/or tumour specific studies would have to be carried out in order to identify the upregulated SR protein(s). The problem with targeting essential cellular proteins such as SR proteins using new drugs is how does one specifically target only tumour cells? Some strategies have been discussed that might prove useful in the future. For example research has been going into investigating the efficacy and safety of viral vectors for treatment of diseases including cancers. Viruses could be engineered to infect only cells expressing tumour specific cell surface markers which would allow targeting of tumour cells specifically. Cervical cancers have a unique possibility in that if therapies could be formulated to be applied and taken up across the plasma membrane of the surface cells then this could theoretically be used to treat cervical cancers. On the other hand the possible benefit of targeting SR proteins is that some of their functions appear redundant, where loss of one protein can be compensated for by another SR protein and in addition as demonstrated by my studies and the investigation by Karni et al. (2007) a complete knockdown in SR protein levels is not necessary to reverse the phenotype.

Recently, there has been some very interesting research into the possibilities of targeting SR proteins as a treatment for some diseases. In my view SR proteins

would provide an excellent target for therapy in for example, cancers where particular SR proteins have been demonstrated to be overexpressed, and other diseases such as HPV and other viral infections where alternative splicing, and therefore SR proteins, is essential for completion of the infectious life cycle. As has been suggested by Karni et al. (2007), complete removal of an SR protein, in the case of cancers, may not be necessary, merely the reduction in expression of the SR protein to near normal levels could be sufficient to reverse the transformed phenotype (Karni et al., 2007). Recent studies by the Tazi group have focused on the use of Indole Derivative Compounds (IDCs) against SR proteins in the treatment of HIV infection. IDCs bind to the RS domain of the target SR protein and prevent the phosphorylation of the RS domain by SR kinases (Bakkour et al., 2007). It is not known exactly what effect this inhibition of phosphorylation is having on the action of the SR proteins. It could be preventing their recognition of ESEs or inhibiting their export functions among others and this will need to be investigated before IDCs could be considered for drug therapies (Keriel et al., 2009).

HIV-1 is another virus, along with HPV, which requires extensive alternative splicing in order to express all viral RNAs and complete its life cycle. Bakkour et al. (2007) have shown that inhibiting SRSF1, which is essential for HIV-1 RNA splicing, prevents progeny virion assembly and release, thereby preventing the spread of virus to other cells in the surrounding area and also the spread of virus between individuals (Bakkour et al., 2007). Bakkour et al. (2007) reported that introduction of IDC16 into cell lines, including primary cell lines, had no negative effects on cellular growth and of 96 cellular targets analysed only 15 displayed any alteration in alterative splicing upon introduction of IDC16 (Bakkour et al.,

2007). This suggests that IDC16 inhibition of SRSF1 activity is not catastrophic for the cell.

Similarly IDC13 and IDC78 have both been shown to both inhibit Murine Leukemia Virus replication, this time in mice (Keriel et al., 2009). Keriel et al. (2009) showed that in Murine Leukaemia Virus infection *in vitro* targeting of an SR protein also prevented virus release and spread (Keriel et al., 2009). Importantly these studies have also suggested that using IDCs does not appear to have overly deleterious effects on the cell, the cell seems able to survive without SRSF1 as does the whole animal after systemic introduction of IDCs by intra-peritoneal injection (Keriel et al., 2009). An exon array showed that out of 6000 murine transcripts only 52 transcripts displayed any significant fold change of between 1.5 and 3 fold difference (Keriel et al., 2009) which equates to roughly 2% of transcripts were affected by SRSF1 loss. This is probably due to the functional redundancy exhibited by the SR protein family; however the virus does not seem able to overcome this loss of SRSF1. Another reason for the seemingly specific effect on viral RNAs is probably due to the fact that during a viral infection the cell is overwhelmed with viral RNAs and the cellular RNA production is either reduced or prevented. Therefore the loss of the splicing factor is not of great consequence for the small number of cellular RNAs still being expressed. Finally the authors hypothesised that as cellular RNAs seem to be able to continue to be expressed, the action of the IDCs is probably not due to any constitutive functions of the SR proteins such as RNA export or initiation of translation, but is likely due to the recognition of exonic splicing enhancers and therefore alternative splicing functions (Keriel et al., 2009). Cellular RNAs may have constitutive exons where the ESEs contain binding sequences for other SR proteins enhancing the functional redundancy if SR proteins in splicing

events. This means that if one SR protein is altered or functionally impaired another SR protein can also bind, although probably with a lower affinity, and still activate splicing of the RNA.

Therefore if these compounds prove to be specific and effective in the clinical setting, and could be formulated to be topically delivered to the cervix, they could prove to be ideal for treating HPV16 induced cervical cancers. The cervix is accessible through methods similar to those used during cervical smear tests and while they are uncomfortable they are far less invasive than surgical procedures and could be more specifically targeted than by systemic methods. If IDCs could be formulated in a way that would allow them to be topically applied and taken up by the actively dividing tumour cells as opposed to the non-dividing granular layer cells in the surrounding epithelium then a tumour specific therapy could be used. For example if they could be delivered by viral vectors that specifically recognise tumour cells or by lipid based mechanisms then topical therapies could be considered.

The targeting of SR proteins for therapy could not only help treat cervical cancers but could also potentially be used as antiviral therapy against both high and low risk mucosal HPV viruses. High risk HPV16 E2 has previously been shown to upregulate a subset of SR proteins, which suggests these proteins are required for completion of the viral life cycle (Mole et al., 2009a, Mole et al., 2009b). Figure 3.7 indicates that low risk viruses may also upregulate certain SR proteins. This suggests low risk viruses also require SR protein functions, whether it is their alternative splicing functions or other functions, in order to complete their life cycle. The targeting of SR proteins could then, theoretically, be applied to all stages of CIN lesion and also genital warts. Therefore for

treatment considerations, targeting SRSF2 should be beneficial for treating HPV16 transformed cells; however one would need to target SRSF3 to treat productive infections. My data has shown that targeting SRSF2 in HPV16-positive tumour cells reverses the transformed phenotype by inducing apoptosis and reducing proliferation, therefore targeting SRSF2 in cancer cells could inhibit tumour growth. However if a productive viral infection is being targeted then SRSF2 is not the best target for therapy. As shown in Figure 4.4, SRSF3 is required to greater extent than SRSF2 for viral RNA processing. Unpublished information from our laboratory indicates that SRSF3 is required for late RNA processing and L1 protein expression (T. Klymenko and SV. Graham, unpublished information). Therefore if viral infections are to be inhibited then SRSF3 is a better therapeutic target than SRSF2. The method of delivery and ease of access to the lesion is the same in both cases.

5.2 Future Directions

In order to fully complete this work and before any therapeutic considerations could be made, the global cellular effects of knocking down SRSF2 would have to be investigated. For example an exon microarray analysis or deep sequencing should be performed to find out which host RNAs are affected by SRSF2 reduction. For example RNA expression in control and SRSF2 knockdown W12GPXY cells should be analysed to determine whether SRSF2 knockdown results in a specific down-regulation of E6/E7 mRNAs and other SRSF2 controlled mRNAs or whether there is a general RNA processing effect after SRSF2 knockdown leading to catastrophic events in the transfected cells. Similar

experiments should be carried out after control and SRSF3 knockdown in W12E cells and both viral and cellular mRNA expression analysed. Unfortunately this experiment was outside the cost and time restraints of this PhD project.

This would be particularly important for developing new avenues to therapy for treating other types of cancer; however one might argue that catastrophic consequences are not as important when targeting cervical tumour cells where the tumour is relatively easy to access. Cervical tumours are one of the more accessible tumours so extremely invasive techniques would not be required to specifically target the tumour. Therefore if topical therapies could be developed as opposed to systemic treatments then perhaps any global cellular effects of targeting SR proteins would not be as important a consideration for cervical cancers as it is for other cancers.

We have been able to demonstrate increased SR protein levels upon transformation in patient tissue, however, the splicing pattern of E6/E7 RNAs would also have to be investigated in high grade lesions from patients. For example FISH assays could be performed in patient tissue to continue these observations. Fluorescently labelled cross splice-junction probes could be used to detect the presence of individual E6 isoforms in patient tissue. If any increases in small isoform production with cervical cancer progression could be replicated in patient tissue it could perhaps lead to new treatment targets for HPV induced cervical cancer. For example if the small isoforms are increased in expression upon transformation of real patient cervical epithelial cells, then perhaps targeting of the individual small isoforms using cross-splice junction siRNAs could potentially inhibit transformation of the cervical epithelial cells.

In addition, siRNA knockdown of individual E6 isoforms would aid our attempts to determine function for the isoforms. Such experiments were initially planned to be carried out, however the struggle to design isoform specific probe and primer sets postponed these experiments as the splice junctions within E6/E7 gene region do not provide good sequences for primer and siRNA design and so even if individual isoforms were knocked down by siRNA targeting, there would be no easy means of quantifying changes in RNA isoform production. In addition, the lack of isoform specific detection at the protein level also hampered these experiments as analysis of knockdown by protein levels would not be possible.

SRSF2 is an RNA-binding protein and could be expected to bind E6/E7 RNAs to protect them from decay. To test this, further experiments should also be completed to confirm the hypothesis that SRSF2 binds the E6/E7 RNAs. I have cloned fragments of the E6/E7 open reading frame ready to carry out EMSAs protein/RNA binding assays however I could not do the experiment due to time constraints.

The transcription assays do not give a clear picture of any effect on transcription initiation from the HPV16 LCR in the presence and absence of SRSF2. However, SRSF2 has been shown previously to regulate transcription elongation. Perhaps a better approach to these experiments would have been to perform nuclear run-on studies. Nuclear run on assays can be used to quantify the transcriptional activity of a gene of interest. Cells would be transfected with siSRSF2 and their nucleus harvested. The nuclei would then be incubated with radioactively labelled UTP and other nucleoside triphosphates. The labelled UTP would then be incorporated into elongating mRNAs. The labelled mRNAs would then be hybridised to gene specific single stranded DNA on a membrane and the

frequency of transcription indicated by the strength of the radioactive label. These assays are more specific and can give a quantification of the frequency of transcription without RNA stability issues being a problem. Transcription assays like the ones carried out in this thesis only give a measure of the total number of transcripts not how frequently they are transcribed. These experiments should be attempted in future work.

While the colony formation assays clearly revealed a difference in anchorage independent growth capability after SRSF2 knockdown the experiment that I carried out is not perfect because, despite transfection efficiencies of around 80% I cannot rule out the possibility that the colonies seen in the SRSF2 knockdown plates could have arisen from cells that were not initially transfected. The assay itself must be carried out over a period of time past the normal effectiveness of transient siRNA experiments or in cells where levels of SRSF2 are permanently reduced through stable shRNA knockdown in every cell. An alternative way to carry out this experiment would have been to use an inducible shRNA-stably transfected cell line where treatment with, for example doxycycline, would switch on expression of the shRNA, so that carrying out the experiment over an extended period of time would not cause any problems.

5.3 Conclusions

The main findings from this PhD thesis are that alternative splicing of HPV16 E6 and E7 RNAs is altered during transformation of cervical epithelial cells. Transformation results in a shift towards the expression of small E6/E7 mRNA isoforms and these isoforms have been shown to have growth promoting effects on cervical epithelial cells. In the transformed cervical epithelial cells SR proteins SRSFs 1-3 are overexpressed. This upregulation was also observed in patient tissue. Overexpression of SRSF2 in cervical epithelial cells is oncogenic and is responsible for stabilising E6/E7 mRNAs in HPV16-positive transformed cervical epithelial cells. Knockdown of SRSF2 in transformed cervical epithelial cells results in a reduced proliferation rate, reduced anchorage independent growth ability, altered cell cycle and an increase in apoptosis in the transfected cells. Knockdown of SRSF2 also results in a decrease in E6/E7 oncoprotein mRNA expression. If topically applied therapies could be formulated that would allow specific targeting of SRSF2 in transformed cervical epithelial cells, then these therapies could inhibit tumour growth. My data suggests that SRSF2 should be considered to be a proto-oncogene. Therefore SRSF2 could prove to be an excellent novel therapy for treatment of HPV16-positive cervical cancers.

6 References

Aasen, T., Hodgins, M. B., Edward, M. & Graham, S. V. (2003). The relationship between connexins, gap junctions, tissue architecture and tumour invasion, as studied in a novel in vitro model of HPV-16-associated cervical cancer progression. *Oncogene* **22**, 7969-80.

Adams, J. M. & Cory, S. (2007). The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* **26**, 1324-37.

Allen-Hoffmann, B. L., Schlosser, S. J., Ivarie, C. A., Sattler, C. A., Meisner, L. F. & O'Connor, S. L. (2000). Normal growth and differentiation in a spontaneously immortalized near-diploid human keratinocyte cell line, NIKS. *J Invest Dermatol* **114**, 444-55.

Ashrafi, G. H., Haghshenas, M. R., Marchetti, B., O'Brien, P. M. & Campo, M. S. (2005). E5 protein of human papillomavirus type 16 selectively downregulates surface HLA class I. *Int J Cancer* **113**, 276-83.

Bakkour, N., Lin, Y. L., Maire, S., Ayadi, L., Mahuteau-Betzer, F., Nguyen, C. H., Mettling, C., Portales, P., Grierson, D., Chabot, B., Jeanteur, P., Branlant, C., Corbeau, P. & Tazi, J. (2007). Small-molecule inhibition of HIV pre-mRNA splicing as a novel antiretroviral therapy to overcome drug resistance. *PLoS Pathog* **3**, 1530-9.

Bechtold, V., Beard, P. & Raj, K. (2003). Human papillomavirus type 16 E2 protein has no effect on transcription from episomal viral DNA. *J Virol* **77**, 2021-8.

Bernard, B. A., Bailly, C., Lenoir, M. C., Darmon, M., Thierry, F. & Yaniv, M. (1989). The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV18 regulatory region in human keratinocytes. *J Virol* **63**, 4317-24.

Bernard, H. U. (2005). The clinical importance of the nomenclature, evolution and taxonomy of human papillomaviruses. *J Clin Virol* **32 Suppl 1**, S1-6.

Bernat, A., Avvakumov, N., Mymryk, J. S. & Banks, L. (2003). Interaction between the HPV E7 oncoprotein and the transcriptional coactivator p300. *Oncogene* **22**, 7871-81.

Boxman, I. L., Russell, A., Mulder, L. H., Bavinck, J. N., ter Schegget, J. & Green, A. (2001). Association between epidermodysplasia verruciformis-associated human papillomavirus DNA in plucked eyebrow hair and solar keratoses. *J Invest Dermatol* **117**, 1108-12.

Boyer, S. N., Wazer, D. E. & Band, V. (1996). E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. *Cancer Res* **56**, 4620-4.

Breitburd, F., Kirnbauer, R., Hubbert, N. L., Nonnenmacher, B., Trin-Dinh-Desmarquet, C., Orth, G., Schiller, J. T. & Lowy, D. R. (1995). Immunization with viruslike particles from cottontail rabbit papillomavirus (CRPV) can protect against experimental CRPV infection. *J Virol* **69**, 3959-63.

Brogna, S. & Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Mol Biol* **16**, 107-13.

Bulk, S., Berkhof, J., Bulkmans, N. W., Zielinski, G. D., Rozendaal, L., van Kemenade, F. J., Snijders, P. J. & Meijer, C. J. (2006). Preferential risk of HPV16 for squamous cell carcinoma and of HPV18 for adenocarcinoma of the cervix compared to women with normal cytology in The Netherlands. *Br J Cancer* **94**, 171-5.

Caceres, J. F., Screaton, G. R. & Krainer, A. R. (1998). A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev* **12**, 55-66.

Chattopadhyay, S., Yamashita, M., Zhang, Y. & Sen, G. C. (2011). The IRF-3/Bax-mediated apoptotic pathway, activated by viral cytoplasmic RNA and DNA, inhibits virus replication. *J Virol* **85**, 3708-16.

Chellappan, S., Kraus, V. B., Kroger, B., Munger, K., Howley, P. M., Phelps, W. C. & Nevins, J. R. (1992). Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product. *Proc Natl Acad Sci U S A* **89**, 4549-53.

Chen, J. J., Reid, C. E., Band, V. & Androphy, E. J. (1995). Interaction of papillomavirus E6 oncoproteins with a putative calcium-binding protein. *Science* **269**, 529-31.

Cheng, H., Dufu, K., Lee, C. S., Hsu, J. L., Dias, A. & Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* **127**, 1389-400.

Cheunim, T., Zhang, J., Milligan, S. G., McPhillips, M. G. & Graham, S. V. (2008). The alternative splicing factor hnRNP A1 is up-regulated during virus-infected epithelial cell differentiation and binds the human papillomavirus type 16 late regulatory element. *Virus Res* **131**, 189-98.

Chiarugi, P. & Giannoni, E. (2008). Anoikis: a necessary death program for anchorage-dependent cells. *Biochem Pharmacol* **76**, 1352-64.

Cho, K. O., Hunt, C. A. & Kennedy, M. B. (1992). The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumor suppressor protein. *Neuron* **9**, 929-42.

Cho, S., Hoang, A., Sinha, R., Zhong, X. Y., Fu, X. D., Krainer, A. R. & Ghosh, G. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc Natl Acad Sci U S A* **108**, 8233-8.

Choo, K. B., Pan, C. C. & Han, S. H. (1987). Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: preferential deletion of the E2 gene and invariable retention of the long control region and the E6/E7 open reading frames. *Virology* **161**, 259-61.

Clower, R. V., Fisk, J. C. & Melendy, T. (2006). Papillomavirus E1 protein binds to and stimulates human topoisomerase I. *J Virol* **80**, 1584-7.

Cooke, C. & Alwine, J. C. (1996). The cap and the 3' splice site similarly affect polyadenylation efficiency. *Mol Cell Biol* **16**, 2579-84.

Crook, T., Tidy, J. A. & Vousden, K. H. (1991). Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation. *Cell* **67**, 547-56.

Cullen, B. R. (2003). Nuclear RNA export. *J Cell Sci* **116**, 587-97.

Cumming, S. A., Chuen-Im, T., Zhang, J. & Graham, S. V. (2009). The RNA stability regulator HuR regulates L1 protein expression in vivo in differentiating cervical epithelial cells. *Virology* **383**, 142-9.

Davy, C. E., Jackson, D. J., Raj, K., Peh, W. L., Southern, S. A., Das, P., Sorathia, R., Laskey, P., Middleton, K., Nakahara, T., Wang, Q., Masterson, P. J., Lambert, P. F., Cuthill, S., Millar, J. B. & Doorbar, J. (2005). Human papillomavirus type 16 E1 E4-induced G2 arrest is associated with cytoplasmic retention of active Cdk1/cyclin B1 complexes. *J Virol* **79**, 3998-4011.

Davy, C. E., Jackson, D. J., Wang, Q., Raj, K., Masterson, P. J., Fenner, N. F., Southern, S., Cuthill, S., Millar, J. B. & Doorbar, J. (2002). Identification

of a G(2) arrest domain in the E1 wedge E4 protein of human papillomavirus type 16. *J Virol* **76**, 9806-18.

de Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U. & zur Hausen, H. (2004). Classification of papillomaviruses. *Virology* **324**, 17-27.

DeFilippis, R. A., Goodwin, E. C., Wu, L. & DiMaio, D. (2003). Endogenous human papillomavirus E6 and E7 proteins differentially regulate proliferation, senescence, and apoptosis in HeLa cervical carcinoma cells. *J Virol* **77**, 1551-63.

Donaldson, M. M., Boner, W. & Morgan, I. M. (2007). TopBP1 regulates human papillomavirus type 16 E2 interaction with chromatin. *J Virol* **81**, 4338-42.

Doorbar, J. (2005). The papillomavirus life cycle. *J Clin Virol* **32 Suppl 1**, S7-15.

Doorbar, J. & Gallimore, P. H. (1987). Identification of proteins encoded by the L1 and L2 open reading frames of human papillomavirus 1a. *J Virol* **61**, 2793-9.

Dow, L. E., Brumby, A. M., Muratore, R., Coombe, M. L., Sedelies, K. A., Trapani, J. A., Russell, S. M., Richardson, H. E. & Humbert, P. O. (2003). hScrib is a functional homologue of the Drosophila tumour suppressor Scribble. *Oncogene* **22**, 9225-30.

Duensing, A., Chin, A., Wang, L., Kuan, S. F. & Duensing, S. (2008). Analysis of centrosome overduplication in correlation to cell division errors in high-risk human papillomavirus (HPV)-associated anal neoplasms. *Virology* **372**, 157-64.

Duensing, S. & Munger, K. (2003). Human papillomavirus type 16 E7 oncoprotein can induce abnormal centrosome duplication through a mechanism independent of inactivation of retinoblastoma protein family members. *J Virol* **77**, 12331-5.

Dyson, N., Howley, P. M., Munger, K. & Harlow, E. (1989). The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science* **243**, 934-7.

Filippova, M., Filippov, V. A., Kagoda, M., Garnett, T., Fodor, N. & Duerksen-Hughes, P. J. (2009). Complexes of human papillomavirus type 16 E6 proteins form pseudo-death-inducing signaling complex structures during tumor necrosis factor-mediated apoptosis. *J Virol* **83**, 210-27.

Filippova, M., Johnson, M. M., Bautista, M., Filippov, V., Fodor, N., Tungteakkhun, S. S., Williams, K. & Duerksen-Hughes, P. J. (2007). The large and small isoforms of human papillomavirus type 16 E6 bind to and differentially affect procaspase 8 stability and activity. *J Virol* **81**, 4116-29.

Filippova, M., Parkhurst, L. & Duerksen-Hughes, P. J. (2004). The human papillomavirus 16 E6 protein binds to Fas-associated death domain and protects cells from Fas-triggered apoptosis. *J Biol Chem* **279**, 25729-44.

Filippova, M., Song, H., Connolly, J. L., Dermody, T. S. & Duerksen-Hughes, P. J. (2002). The human papillomavirus 16 E6 protein binds to tumor necrosis factor (TNF) R1 and protects cells from TNF-induced apoptosis. *J Biol Chem* **277**, 21730-9.

Fischer, D. C., Noack, K., Runnebaum, I. B., Watermann, D. O., Kieback, D. G., Stamm, S. & Stickeler, E. (2004). Expression of splicing factors in human ovarian cancer. *Oncol Rep* **11**, 1085-90.

Flores, E. R., Allen-Hoffmann, B. L., Lee, D., Sattler, C. A. & Lambert, P. F. (1999). Establishment of the human papillomavirus type 16 (HPV-16) life cycle in an immortalized human foreskin keratinocyte cell line. *Virology* **262**, 344-54.

Frattini, M. G. & Laimins, L. A. (1994). Binding of the human papillomavirus E1

origin-recognition protein is regulated through complex formation with the E2 enhancer-binding protein. *Proc Natl Acad Sci U S A* **91**, 12398-402.
Fuchs, E. (1990). Epidermal differentiation: the bare essentials. *J Cell Biol* **111**, 2807-14.
Fujita, M., Kiyono, T., Hayashi, Y. & Ishibashi, M. (1996). hCDC47, a human member of the MCM family. Dissociation of the nucleus-bound form during S phase. *J Biol Chem* **271**, 4349-54.
Funk, J. O., Waga, S., Harry, J. B., Espling, E., Stillman, B. & Galloway, D. A. (1997). Inhibition of CDK activity and PCNA-dependent DNA replication by p21 is blocked by interaction with the HPV-16 E7 oncoprotein. *Genes Dev* **11**, 2090-100.
Gardiol, D., Kuhne, C., Glaunsinger, B., Lee, S. S., Javier, R. & Banks, L. (1999). Oncogenic human papillomavirus E6 proteins target the discs large tumour suppressor for proteasome-mediated degradation. *Oncogene* **18**, 5487-96.
Ghittoni, R., Accardi, R., Hasan, U., Gheit, T., Sylla, B. & Tommasino, M. (2010). The biological properties of E6 and E7 oncoproteins from human papillomaviruses. *Virus Genes* **40**, 1-13.
Glaunsinger, B. A., Lee, S. S., Thomas, M., Banks, L. & Javier, R. (2000). Interactions of the PDZ-protein MAGI-1 with adenovirus E4-ORF1 and high-risk papillomavirus E6 oncoproteins. *Oncogene* **19**, 5270-80.
Goodman, R. H. & Smolik, S. (2000). CBP/p300 in cell growth, transformation, and development. *Genes Dev* **14**, 1553-77.
Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. *Rna* **6**, 1197-211.
Gray, E., Pett, M. R., Ward, D., Winder, D. M., Stanley, M. A., Roberts, I., Scarpini, C. G. & Coleman, N. (2010). In vitro progression of human papillomavirus 16 episome-associated cervical neoplasia displays fundamental similarities to integrant-associated carcinogenesis. *Cancer Res* **70**, 4081-91.
Greenstone, H. L., Nieland, J. D., de Visser, K. E., De Bruijn, M. L., Kirnbauer, R., Roden, R. B., Lowy, D. R., Kast, W. M. & Schiller, J. T. (1998). Chimeric papillomavirus virus-like particles elicit antitumor immunity against the E7 oncoprotein in an HPV16 tumor model. *Proc Natl Acad Sci U S A* **95**, 1800-5.
Grm, H. S. & Banks, L. (2004). Degradation of hDlg and MAGIs by human papillomavirus E6 is E6-AP-independent. *J Gen Virol* **85**, 2815-9.
Gross-Mesilaty, S., Reinstein, E., Bercovich, B., Tobias, K. E., Schwartz, A. L., Kahana, C. & Ciechanover, A. (1998). Basal and human papillomavirus E6 oncoprotein-induced degradation of Myc proteins by the ubiquitin pathway. *Proc Natl Acad Sci U S A* **95**, 8058-63.
Hamm, J. & Mattaj, I. W. (1990). Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* **63**, 109-18.
Harwood, C. A., Surentheran, T., McGregor, J. M., Spink, P. J., Leigh, I. M., Breuer, J. & Proby, C. M. (2000). Human papillomavirus infection and non-melanoma skin cancer in immunosuppressed and immunocompetent individuals. *J Med Virol* **61**, 289-97.
Hasan, U. A., Bates, E., Takeshita, F., Biliato, A., Accardi, R., Bouvard, V., Mansour, M., Vincent, I., Gissmann, L., Iftner, T., Sideri, M., Stubenrauch, F. & Tommasino, M. (2007). TLR9 expression and function is abolished by the cervical cancer-associated human papillomavirus type 16. *J Immunol* **178**, 3186-97.
He, X., Arslan, A. D., Pool, M. D., Ho, T. T., Darcy, K. M., Coon, J. S. & Beck, W. T. (2011). Knockdown of splicing factor SRp20 causes apoptosis in ovarian

cancer cells and its expression is associated with malignancy of epithelial ovarian cancer. *Oncogene* **30**, 356-65.
Hengstermann, A., Linares, L. K., Ciechanover, A., Whitaker, N. J. & Scheffner, M. (2001). Complete switch from Mdm2 to human papillomavirus E6-mediated degradation of p53 in cervical cancer cells. *Proc Natl Acad Sci U S A* **98**, 1218-23.
Hirasawa, A., Saito-Ohara, F., Inoue, J., Aoki, D., Susumu, N., Yokoyama, T., Nozawa, S., Inazawa, J. & Imoto, I. (2003). Association of 17q21-q24 gain in ovarian clear cell adenocarcinomas with poor prognosis and identification of PPM1D and APPBP2 as likely amplification targets. *Clin Cancer Res* **9**, 1995-2004.
Hocine, S., Singer, R. H. & Grunwald, D. (2010). RNA processing and export. *Cold Spring Harb Perspect Biol* **2**, a000752.
Huang, Y., Gattoni, R., Stevenin, J. & Steitz, J. A. (2003). SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell* **11**, 837-43.
Huang, Y. & Steitz, J. A. (2001). Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol Cell* **7**, 899-905.
Huang, Y., Yario, T. A. & Steitz, J. A. (2004). A molecular link between SR protein dephosphorylation and mRNA export. *Proc Natl Acad Sci U S A* **101**, 9666-70.
Hughes, F. J. & Romanos, M. A. (1993). E1 protein of human papillomavirus is a DNA helicase/ATPase. *Nucleic Acids Res* **21**, 5817-23.
Huh, K. W., DeMasi, J., Ogawa, H., Nakatani, Y., Howley, P. M. & Munger, K. (2005). Association of the human papillomavirus type 16 E7 oncoprotein with the 600-kDa retinoblastoma protein-associated factor, p600. *Proc Natl Acad Sci U S A* **102**, 11492-7.
Huibregtse, J. M., Scheffner, M. & Howley, P. M. (1991). A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *Embo J* **10**, 4129-35.
Humbert, P., Russell, S. & Richardson, H. (2003). Dlg, Scribble and Lgl in cell polarity, cell proliferation and cancer. *Bioessays* **25**, 542-53.
Iftner, T., Elbel, M., Schopp, B., Hiller, T., Loizou, J. I., Caldecott, K. W. & Stubenrauch, F. (2002). Interference of papillomavirus E6 protein with single-strand break repair by interaction with XRCC1. *Embo J* **21**, 4741-8.
Jeon, S., Allen-Hoffmann, B. L. & Lambert, P. F. (1995). Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J Virol* **69**, 2989-97.
Jeon, S. & Lambert, P. F. (1995). Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. *Proc Natl Acad Sci U S A* **92**, 1654-8.
Jia, R., Liu, X., Tao, M., Kruhlak, M., Guo, M., Meyers, C., Baker, C. C. & Zheng, Z. M. (2009). Control of the papillomavirus early-to-late switch by differentially expressed SRp20. *J Virol* **83**, 167-80.
Jones, D. L., Alani, R. M. & Munger, K. (1997). The human papillomavirus E7 oncoprotein can uncouple cellular differentiation and proliferation in human keratinocytes by abrogating p21Cip1-mediated inhibition of cdk2. *Genes Dev* **11**, 2101-11.
Kaikkonen, M. U., Lam, M. T. & Glass, C. K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc Res* **90**, 430-40.
Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W. & Waldman, F. M. (1994). Detection

and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci U S A* **91**, 2156-60.

Karni, R., de Stanchina, E., Lowe, S. W., Sinha, R., Mu, D. & Krainer, A. R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* **14**, 185-93.

Keriel, A., Mahuteau-Betzer, F., Jacquet, C., Plays, M., Grierson, D., Sitbon, M. & Tazi, J. (2009). Protection against retrovirus pathogenesis by SR protein inhibitors. *PLoS One* **4**, e4533.

Kessis, T. D., Slebos, R. J., Nelson, W. G., Kastan, M. B., Plunkett, B. S., Han, S. M., Lorincz, A. T., Hedrick, L. & Cho, K. R. (1993). Human papillomavirus 16 E6 expression disrupts the p53-mediated cellular response to DNA damage. *Proc Natl Acad Sci U S A* **90**, 3988-92.

Kines, R. C., Thompson, C. D., Lowy, D. R., Schiller, J. T. & Day, P. M. (2009). The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding. *Proc Natl Acad Sci U S A* **106**, 20458-63.

Kirnbauer, R., Chandrachud, L. M., O'Neil, B. W., Wagner, E. R., Grindlay, G. J., Armstrong, A., McGarvie, G. M., Schiller, J. T., Lowy, D. R. & Campo, M. S. (1996). Virus-like particles of bovine papillomavirus type 4 in prophylactic and therapeutic immunization. *Virology* **219**, 37-44.

Kiyono, T., Hiraiwa, A., Fujita, M., Hayashi, Y., Akiyama, T. & Ishibashi, M. (1997). Binding of high-risk human papillomavirus E6 oncoproteins to the human homologue of the Drosophila discs large tumor suppressor protein. *Proc Natl Acad Sci U S A* **94**, 11612-6.

Korn, W. M., Oide Weghuis, D. E., Suijkerbuijk, R. F., Schmidt, U., Otto, T., du Manoir, S., Geurts van Kessel, A., Harstrick, A., Seeber, S. & Becher, R. (1996). Detection of chromosomal DNA gains and losses in testicular germ cell tumors by comparative genomic hybridization. *Genes Chromosomes Cancer* **17**, 78-87.

Kosel, S., Burggraf, S., Engelhardt, W. & Olgemoller, B. (2007). Increased levels of HPV16 E6*I transcripts in high-grade cervical cytology and histology (CIN II+) detected by rapid real-time RT-PCR amplification. *Cytopathology* **18**, 290-9.

Krajewski, S., Krajewska, M. & Reed, J. C. (1996). Immunohistochemical analysis of in vivo patterns of Bak expression, a proapoptotic member of the Bcl-2 protein family. *Cancer Res* **56**, 2849-55.

Kuhn, U., Gundel, M., Knoth, A., Kerwitz, Y., Rudel, S. & Wahle, E. (2009). Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. *J Biol Chem* **284**, 22803-14.

Kukimoto, I., Aihara, S., Yoshiike, K. & Kanda, T. (1998). Human papillomavirus oncoprotein E6 binds to the C-terminal region of human minichromosome maintenance 7 protein. *Biochem Biophys Res Commun* **249**, 258-62.

Laura, R. P., Ross, S., Koeppen, H. & Lasky, L. A. (2002). MAGI-1: a widely expressed, alternatively spliced tight junction protein. *Exp Cell Res* **275**, 155-70.

Lee, D., Lee, B., Kim, J., Kim, D. W. & Choe, J. (2000). cAMP response element-binding protein-binding protein binds to human papillomavirus E2 protein and activates E2-dependent transcription. *J Biol Chem* **275**, 7045-51.

Lee, S. S., Glaunsinger, B., Mantovani, F., Banks, L. & Javier, R. T. (2000). Multi-PDZ domain protein MUPP1 is a cellular target for both adenovirus E4-ORF1 and high-risk papillomavirus type 18 E6 oncoproteins. *J Virol* **74**, 9680-93.

Lemaire, R., Prasad, J., Kashima, T., Gustafson, J., Manley, J. L. & Lafyatis, R. (2002). Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins. *Genes Dev* **16**, 594-607.

Li, X., Wang, J. & Manley, J. L. (2005). Loss of splicing factor ASF/SF2 induces G2 cell cycle arrest and apoptosis, but inhibits internucleosomal DNA fragmentation. *Genes Dev* **19**, 2705-14.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S. & Fu, X. D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* **15**, 819-26.

Long, J. C. & Caceres, J. F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* **417**, 15-27.

Loo, Y. M. & Melendy, T. (2004). Recruitment of replication protein A by the papillomavirus E1 protein and modulation by single-stranded DNA. *J Virol* **78**, 1605-15.

Loomis, R. J., Naoe, Y., Parker, J. B., Savic, V., Bozovsky, M. R., Macfarlan, T., Manley, J. L. & Chakravarti, D. (2009). Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. *Mol Cell* **33**, 450-61.

Lu, Z., Hu, X., Li, Y., Zheng, L., Zhou, Y., Jiang, H., Ning, T., Basang, Z., Zhang, C. & Ke, Y. (2004). Human papillomavirus 16 E6 oncoprotein interferences with insulin signaling pathway by binding to tuberin. *J Biol Chem* **279**, 35664-70.

Luo, M. L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M. & Reed, R. (2001). Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413**, 644-7.

Maglennon, G. A., McIntosh, P. & Doorbar, J. (2011). Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression. *Virology* **414**, 153-63.

Manley, J. L. & Krainer, A. R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev* **24**, 1073-4.

McLaughlin-Drubin, M. E., Huh, K. W. & Munger, K. (2008). Human papillomavirus type 16 E7 oncoprotein associates with E2F6. *J Virol* **82**, 8695-705.

McPhillips, M. G., Oliveira, J. G., Spindler, J. E., Mitra, R. & McBride, A. A. (2006). Brd4 is required for e2-mediated transcriptional activation but not genome partitioning of all papillomaviruses. *J Virol* **80**, 9530-43.

McPhillips, M. G., Veerapraditsin, T., Cumming, S. A., Karali, D., Milligan, S. G., Boner, W., Morgan, I. M. & Graham, S. V. (2004). SF2/ASF binds the human papillomavirus type 16 late RNA control element and is regulated during differentiation of virus-infected epithelial cells. *J Virol* **78**, 10598-605.

Melillo, R. M., Helin, K., Lowy, D. R. & Schiller, J. T. (1994). Positive and negative regulation of cell proliferation by E2F-1: influence of protein level and human papillomavirus oncoproteins. *Mol Cell Biol* **14**, 8241-9.

Michlewski, G., Sanford, J. R. & Caceres, J. F. (2008). The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Mol Cell* **30**, 179-89.

Middleton, K., Peh, W., Southern, S., Griffin, H., Sotlar, K., Nakahara, T., El-Sherif, A., Morris, L., Seth, R., Hibma, M., Jenkins, D., Lambert, P., Coleman, N. & Doorbar, J. (2003). Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers. *J Virol* **77**, 10186-201.

Milligan, S. G., Veerapraditsin, T., Ahamet, B., Mole, S. & Graham, S. V. (2007). Analysis of novel human papillomavirus type 16 late mRNAs in differentiated W12 cervical epithelial cells. *Virology* **360**, 172-81.

Mole, S., McFarlane, M., Chuen-Im, T., Milligan, S. G., Millan, D. & Graham, S. V. (2009). RNA splicing factors regulated by HPV16 during cervical tumour progression. *J Pathol* **219**, 383-91.

Mole, S., Milligan, S. G. & Graham, S. V. (2009). Human papillomavirus type 16 E2 protein transcriptionally activates the promoter of a key cellular splicing factor, SF2/ASF. *J Virol* **83**, 357-67.

Moll, U. M., Wolff, S., Speidel, D. & Deppert, W. (2005). Transcription-independent pro-apoptotic functions of p53. *Curr Opin Cell Biol* **17**, 631-6.

Moore, M. J. & Proudfoot, N. J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688-700.

Nakagawa, S. & Huibregtse, J. M. (2000). Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. *Mol Cell Biol* **20**, 8244-53.

Nguyen, C. L., Eichwald, C., Nibert, M. L. & Munger, K. (2007). Human papillomavirus type 16 E7 oncoprotein associates with the centrosomal component gamma-tubulin. *J Virol* **81**, 13533-43.

Nguyen, C. L. & Munger, K. (2008). Direct association of the HPV16 E7 oncoprotein with cyclin A/CDK2 and cyclin E/CDK2 complexes. *Virology* **380**, 21-5.

Nguyen, C. L. & Munger, K. (2009). Human papillomavirus E7 protein deregulates mitosis via an association with nuclear mitotic apparatus protein 1. *J Virol* **83**, 1700-7.

Nicolaides, L., Davy, C., Raj, K., Kranjec, C., Banks, L. & Doorbar, J. (2011). Stabilization of HPV16 E6 protein by PDZ proteins, and potential implications for genome maintenance. *Virology* **414**, 137-45.

Nomine, Y., Masson, M., Charbonnier, S., Zanier, K., Ristriani, T., Deryckere, F., Sibler, A.P., Desplancq, D., Atkinson, R.A., Weiss, E., Orfanoudakis, G., Kieffer,B. & Trave, G. (2006). Structural and functional analysis of E6 oncoprotein: insights in the molecular pathways of human papillomavirus-mediated pathogenesis. *Mol Cell* **21**, 665-678.

Ohlenschlager, O., Seiboth, T., Zengerling, H., Briese, L., Marchanka, A., Ramachandran, R., Baum, M., Korbas, M., Meyer-Klaucke, W., Durst, M., Gorlach, M. (2006). Solution structure of the patially folded high-risk human papilloma virus 45 oncoprotein E7. *Oncogene* **25,** 5953-5959.

Ola, M. S., Nawaz, M. & Ahsan, H. (2011). Role of Bcl-2 family proteins and caspases in the regulation of apoptosis. *Mol Cell Biochem* **351**, 41-58.

Ottinger, M., Smith, J. A., Schweiger, M. R., Robbins, D., Powell, M. L., You, J. & Howley, P. M. (2009). Cell-type specific transcriptional activities among different papillomavirus long control regions and their regulation by E2. *Virology* **395**, 161-71.

Ozbun, M. A. (2002). Human papillomavirus type 31b infection of human keratinocytes and the onset of early transcription. *J Virol* **76**, 11291-300.

Parish, J. L., Bean, A. M., Park, R. B. & Androphy, E. J. (2006). ChlR1 is required for loading papillomavirus E2 onto mitotic chromosomes and viral genome maintenance. *Mol Cell* **24**, 867-76.

Parish, J. L., Kowalczyk, A., Chen, H. T., Roeder, G. E., Sessions, R., Buckle, M. & Gaston, K. (2006). E2 proteins from high- and low-risk human papillomavirus types differ in their ability to bind p53 and induce apoptotic cell death. *J Virol* **80**, 4580-90.

Park, P., Copeland, W., Yang, L., Wang, T., Botchan, M. R. & Mohr, I. J. (1994). The cellular DNA polymerase alpha-primase is required for papillomavirus DNA replication and associates with the viral E1 helicase. *Proc Natl Acad Sci U S A* **91**, 8700-4.

Phelps, W. C. & Howley, P. M. (1987). Transcriptional trans-activation by the human papillomavirus type 16 E2 gene product. *J Virol* **61**, 1630-8.

Pierrat, O. A., Mikitova, V., Bush, M. S., Browning, K. S. & Doonan, J. H. (2007). Control of protein translation by phosphorylation of the mRNA 5'-cap-binding complex. *Biochem Soc Trans* **35**, 1634-7.

Pim, D., Collins, M. & Banks, L. (1992). Human papillomavirus type 16 E5 gene stimulates the transforming activity of the epidermal growth factor receptor. *Oncogene* **7**, 27-32.

Pim, D., Massimi, P. & Banks, L. (1997). Alternatively spliced HPV-18 E6* protein inhibits E6 mediated degradation of p53 and suppresses transformed cell growth. *Oncogene* **15**, 257-64.

Pucci, B., Kasten, M. & Giordano, A. (2000). Cell cycle and apoptosis. *Neoplasia* **2**, 291-9.

Qian, W., Iqbal, K., Grundke-Iqbal, I., Gong, C. X. & Liu, F. (2011). Splicing factor SC35 promotes tau expression through stabilization of its mRNA. *FEBS Lett* **585**, 875-80.

Ried, T., Petersen, I., Holtgreve-Grez, H., Speicher, M. R., Schrock, E., du Manoir, S. & Cremer, T. (1994). Mapping of multiple DNA gains and losses in primary small cell lung carcinomas by comparative genomic hybridization. *Cancer Res* **54**, 1801-6.

Robberson, B. L., Cote, G. J. & Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* **10**, 84-94.

Rodriguez, M. I., Finbow, M. E. & Alonso, A. (2000). Binding of human papillomavirus 16 E5 to the 16 kDa subunit c (proteolipid) of the vacuolar H+-ATPase can be dissociated from the E5-mediated epidermal growth factor receptor overactivation. *Oncogene* **19**, 3727-32.

Ronco, L. V., Karpova, A. Y., Vidal, M. & Howley, P. M. (1998). Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity. *Genes Dev* **12**, 2061-72.

Rosenberger, S., De-Castro Arce, J., Langbein, L., Steenbergen, R. D. & Rosl, F. (2010). Alternative splicing of human papillomavirus type-16 E6/E6* early mRNA is coupled to EGF signaling via Erk1/2 activation. *Proc Natl Acad Sci U S A* **107**, 7006-11.

Sanford, J. R., Gray, N. K., Beckmann, K. & Caceres, J. F. (2004). A novel role for shuttling SR proteins in mRNA translation. *Genes Dev* **18**, 755-68.

Scheffner, M., Huibregtse, J. M., Vierstra, R. D. & Howley, P. M. (1993). The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* **75**, 495-505.

Schmitt, A., Rochat, A., Zeltner, R., Borenstein, L., Barrandon, Y., Wettstein, F. O. & Iftner, T. (1996). The primary target cells of the high-risk cottontail rabbit papillomavirus colocalize with hair follicle stem cells. *J Virol* **70**, 1912-22.

Schuler, M. & Green, D. R. (2005). Transcription, apoptosis and p53: catch-22. *Trends Genet* **21**, 182-7.

Sedman, S. A., Barbosa, M. S., Vass, W. C., Hubbert, N. L., Haas, J. A., Lowy, D. R. & Schiller, J. T. (1991). The full-length E6 protein of human papillomavirus type 16 has transforming and trans-activating activities and cooperates with E7 to immortalize keratinocytes in culture. *J Virol*

65, 4860-6.
Selinka, H. C., Giroglou, T. & Sapp, M. (2002). Analysis of the infectious entry pathway of human papillomavirus type 33 pseudovirions. *Virology* **299**, 279-287.
Shally, M., Alloul, N., Jackman, A., Muller, M., Gissmann, L. & Sherman, L. (1996). The E6 variant proteins E6I-E6IV of human papillomavirus 16: expression in cell free systems and bacteria and study of their interaction with p53. *Virus Res* **42**, 81-96.
Sheets, M. D. & Wickens, M. (1989). Two phases in the addition of a poly(A) tail. *Genes Dev* **3**, 1401-12.
Shen, H., Kan, J. L. & Green, M. R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell* **13**, 367-76.
Shepard, P. J. & Hertel, K. J. (2009). The SR protein family. *Genome Biol* **10**, 242.
Shimotohno, K., Kodama, Y., Hashimoto, J. & Miura, K. I. (1977). Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. *Proc Natl Acad Sci U S A* **74**, 2734-8.
Shirasawa, H., Jin, M. H., Shimizu, K., Akutsu, N., Shino, Y. & Simizu, B. (1994). Transcription-modulatory activity of full-length E6 and E6*I proteins of human papillomavirus type 16. *Virology* **203**, 36-42.
Smotkin, D., Prokoph, H. & Wettstein, F. O. (1989). Oncogenic and nononcogenic human genital papillomaviruses generate the E7 mRNA by different mechanisms. *J Virol* **63**, 1441-7.
Smotkin, D. & Wettstein, F. O. (1986). Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc Natl Acad Sci U S A* **83**, 4680-4.
Somberg, M. & Schwartz, S. (2010). Multiple ASF/SF2 sites in the human papillomavirus type 16 (HPV-16) E4-coding region promote splicing to the most commonly used 3'-splice site on the HPV-16 genome. *J Virol* **84**, 8219-30.
Soucek, L. & Evan, G. I. (2010). The ups and downs of Myc biology. *Curr Opin Genet Dev* **20**, 91-5.
Srivenugopal, K. S. & Ali-Osman, F. (2002). The DNA repair protein, O(6)-methylguanine-DNA methyltransferase is a proteolytic target for the E6 human papillomavirus oncoprotein. *Oncogene* **21**, 5940-5.
Stacey, S. N., Jordan, D., Snijders, P. J., Mackett, M., Walboomers, J. M. & Arrand, J. R. (1995). Translation of the human papillomavirus type 16 E7 oncoprotein from bicistronic mRNA is independent of splicing events within the E6 open reading frame. *J Virol* **69**, 7023-31.
Staknis, D. & Reed, R. (1994). SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol Cell Biol* **14**, 7670-82.
Stanley, M. (2008). Immunobiology of HPV and HPV vaccines. *Gynecol Oncol* **109**, S15-21.
Stanley, M. A. (2001). Immunobiology of papillomavirus infections. *J Reprod Immunol* **52**, 45-59.
Stanley, M. A., Browne, H. M., Appleby, M. & Minson, A. C. (1989). Properties of a non-tumorigenic human cervical keratinocyte cell line. *Int J Cancer* **43**, 672-6.
Sterling, J. C. (2005). Human papillomaviruses and skin cancer. *J Clin Virol* **32 Suppl 1**, S67-71.

Sterner, J. M., Dew-Knight, S., Musahl, C., Kornbluth, S. & Horowitz, J. M. (1998). Negative regulation of DNA replication by the retinoblastoma protein is mediated by its association with MCM7. *Mol Cell Biol* **18**, 2748-57.

Stickeler, E., Kittrell, F., Medina, D. & Berget, S. M. (1999). Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. *Oncogene* **18**, 3574-82.

Straight, S. W., Hinkle, P. M., Jewers, R. J. & McCance, D. J. (1993). The E5 oncoprotein of human papillomavirus type 16 transforms fibroblasts and effects the downregulation of the epidermal growth factor receptor in keratinocytes. *J Virol* **67**, 4521-32.

Stubenrauch, F. & Laimins, L. A. (1999). Human papillomavirus life cycle: active and latent phases. *Semin Cancer Biol* **9**, 379-86.

Sun, A., Bagella, L., Tutton, S., Romano, G. & Giordano, A. (2007). From G0 to S phase: a view of the roles played by the retinoblastoma (Rb) family members in the Rb-E2F pathway. *J Cell Biochem* **102**, 1400-4.

Tan, S. H., Gloss, B. & Bernard, H. U. (1992). During negative regulation of the human papillomavirus-16 E6 promoter, the viral E2 protein can displace Sp1 from a proximal promoter element. *Nucleic Acids Res* **20**, 251-6.

Tan, S. H., Leong, L. E., Walker, P. A. & Bernard, H. U. (1994). The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. *J Virol* **68**, 6411-20.

Tang, S., Tao, M., McCoy, J. P., Jr. & Zheng, Z. M. (2006). The E7 oncoprotein is translated from spliced E6*I transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation. *J Virol* **80**, 4249-63.

Thomas, M. & Banks, L. (1998). Inhibition of Bak-induced apoptosis by HPV-18 E6. *Oncogene* **17**, 2943-54.

Thomas, M., Laura, R., Hepner, K., Guccione, E., Sawyers, C., Lasky, L. & Banks, L. (2002). Oncogenic human papillomavirus E6 proteins target the MAGI-2 and MAGI-3 proteins for degradation. *Oncogene* **21**, 5088-96.

Thomas, M., Massimi, P., Navarro, C., Borg, J. P. & Banks, L. (2005). The hScrib/Dlg apico-basal control complex is differentially targeted by HPV-16 and HPV-18 E6 proteins. *Oncogene* **24**, 6222-30.

Tong, X. & Howley, P. M. (1997). The bovine papillomavirus E6 oncoprotein interacts with paxillin and disrupts the actin cytoskeleton. *Proc Natl Acad Sci U S A* **94**, 4412-7.

Vogelstein, B., Lane, D. & Levine, A. J. (2000). Surfing the p53 network. *Nature* **408**, 307-10.

Wang, Q., Griffin, H., Southern, S., Jackson, D., Martin, A., McIntosh, P., Davy, C., Masterson, P. J., Walker, P. A., Laskey, P., Omary, M. B. & Doorbar, J. (2004). Functional analysis of the human papillomavirus type 16 E1=E4 protein provides a mechanism for in vivo and in vitro keratin filament reorganization. *J Virol* **78**, 821-33.

Ward, A. J. & Cooper, T. A. (2010). The pathobiology of splicing. *J Pathol* **220**, 152-63.

Wathelet, M. G., Lin, C. H., Parekh, B. S., Ronco, L. V., Howley, P. M. & Maniatis, T. (1998). Virus infection induces the assembly of coordinately activated transcription factors on the IFN-beta enhancer in vivo. *Mol Cell* **1**, 507-18.

Wiegand, H. L., Coburn, G. A., Zeng, Y., Kang, Y., Bogerd, H. P. & Cullen, B. R. (2002). Formation of Tap/NXT1 heterodimers activates Tap-dependent

nuclear mRNA export by enhancing recruitment to nuclear pore complexes. *Mol Cell Biol* **22**, 245-56.

Woods, D. F. & Bryant, P. J. (1991). The discs-large tumor suppressor gene of Drosophila encodes a guanylate kinase homolog localized at septate junctions. *Cell* **66**, 451-64.

Woods, D. F., Hough, C., Peel, D., Callaini, G. & Bryant, P. J. (1996). Dlg protein is required for junction structure, cell polarity, and proliferation control in Drosophila epithelia. *J Cell Biol* **134**, 1469-82.

Wu, J. Y. & Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061-70.

Xiao, R., Sun, Y., Ding, J. H., Lin, S., Rose, D. W., Rosenfeld, M. G., Fu, X. D. & Li, X. (2007). Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol Cell Biol* **27**, 5393-402.

Yang, Y. C., Spalholz, B. A., Rabson, M. S. & Howley, P. M. (1985). Dissociation of transforming and trans-activation functions for bovine papillomavirus type 1. *Nature* **318**, 575-7.

Yee, C., Krishnan-Hewlett, I., Baker, C. C., Schlegel, R. & Howley, P. M. (1985). Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. *Am J Pathol* **119**, 361-6.

You, J., Croyle, J. L., Nishimura, A., Ozato, K. & Howley, P. M. (2004). Interaction of the bovine papillomavirus E2 protein with Brd4 tethers the viral DNA to host mitotic chromosomes. *Cell* **117**, 349-60.

Zerfass-Thome, K., Zwerschke, W., Mannhardt, B., Tindle, R., Botz, J. W. & Jansen-Durr, P. (1996). Inactivation of the cdk inhibitor p27KIP1 by the human papillomavirus type 16 E7 oncoprotein. *Oncogene* **13**, 2323-30.

Zhang, B., Li, P., Wang, E., Brahmi, Z., Dunn, K. W., Blum, J. S. & Roman, A. (2003). The E5 protein of human papillomavirus type 16 perturbs MHC class II antigen maturation in human foreskin keratinocytes treated with interferon-gamma. *Virology* **310**, 100-8.

Zhang, X., Virtanen, A. & Kleiman, F. E. (2010). To polyadenylate or to deadenylate: that is the question. *Cell Cycle* **9**, 4437-49.

Zhang, Z. & Krainer, A. R. (2004). Involvement of SR proteins in mRNA surveillance. *Mol Cell* **16**, 597-607.

Zhou, Z., Luo, M. J., Straesser, K., Katahira, J., Hurt, E. & Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**, 401-5.

Zimmermann, H., Degenkolbe, R., Bernard, H. U. & O'Connor, M. J. (1999). The human papillomavirus type 16 E6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator CBP/p300. *J Virol* **73**, 6209-19.

Zorio, D. A. & Bentley, D. L. (2004). The link between mRNA processing and transcription: communication works both ways. *Exp Cell Res* **296**, 91-7.

7 Appendix 1

8 Appendix 2