# Investigating the Expression of SYCP2 in HPV associated Cancers

**A thesis submitted to the University of Kent for the degree of**

**M.Sc. by Research in Cell Biology to the Faculty of Science,**

**School of Biosciences**

**The University of Kent**

**2021**

Aroub Yousef I Almubarak

Supervisors:

Dr. Tim Fenton, Dr. Peter Ellis

# Declaration

No part of this thesis has been submitted in support of an application for any degree or qualification of the University of Kent or any other University or institute of learning.

Aroub Yousef I Almubarak
January 2021

# Acknowledgments

I would like to thank my supervisors Dr. Tim Fenton and Dr. Peter Ellis for their kind guidance, the team in the Fenton group and Ellis group for their help with my experiments. And Lastly my family for their never ending support.

# Content:


DECLARATION 2

ACKNOWLEDGMENTS 3

ABSTRACT 9

1 INTRODUCTION 10

1.1 Cancer Biology

1.1.1 Cancer and epigenetics 13

1.2 Cancer Testis Antigens

1.2.1 Expression pattern and therapeutic potential of cancer testis antigens 15

1.3 Human Papillomavirus

1.3.1 HPV structure and genome organisation

1.3.2 The HPV life cycle

1.3.3 HPV induced Oncogenesis 25

1.4 Synaptonemal Complex

1.4.1 The Synaptonemal Complex role in DNA replication

1.4.2 The importance of the Synaptonemal Complex for chromosome alignment during meiosis

1.4.3 The Synaptonemal Complex's role in suppressing inter-sister recombination during meiosis

1.4.4 The Synaptonemal Complex gene SYCP2 expression in cancer 35

2. AIMS 37

3. METHODOLOGY 38

3.1 Bioinformatics:

3.2 Tissue Culture:

3.2.1 Cell lines 38

3.3 Protein Analysis:

3.3.1 Protein Extraction

3.3.2 Protein Quantification

3.3.3 Samples prep

3.3.4 SDS PAGE

3.3.5 Western Blotting 41

3.4 Gene Cloning

3.4.1 Bacterial Culture + DNA Purification

3.4.2 Designing Primers

3.4.3 Polymerase Chain Reaction PCR

3.4.4 PCR Product Digestion + Ligation

3.4.5 Plasmid Construct and Bacterial Transformation

3.5 Immunohistochemistry (IHC) 46

4 RESULTS 48

4.1 Bioinformatic Analysis 48

4.2 Western Blots

4.2.1 Optimising Western blotting 53

4.3 Immunofluorescence Microscopy 58

4.4 Cloning Gene of interest into Plasmid Construct 59

5 DISCUSSION: 62

5.1 Assessment of SYCP2 Protein Expression

5.1.1 Human papillomavirus regulation of gene expression effect on host cell genes. 65

5.2 SYCP2 possible role oncogenic molecular events

5.2.1 Alternative Lengthening of Telomeres

5.2.2 What wrongful expression of SYCP2 means to oncogenic progression 67

5.3 Conclusions and Future Prospects 68

6 REFERENCES: 69

# List of Figures:

Figure 1 Epigenetic regulation of CTA expression. 17
Figure 2 Figure shown is the genome of circular double stranded DNA of Human Papillomavirus. Photo Credit: McBride, Alison A. “Mechanisms and strategies of papillomavirus replication.” Biological chemistry vol. 398,8 (2017): 919-927. doi:10.1515/hsz-2017-0113 (McBride, 2017a) 20
Figure 3 The different stages of the papillomavirus life cycle. 24
Figure 4 Model of the elements of the synaptonemal complex. Photo Credit: Molecular Biology of the Cell by Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter, Peter New York and London: Garland Science; c2002 30
Figure 5 Model for role of axial element in promoting interhomolog recombination bias proposed by Li, Xin Chenglin et al. “Genetic evidence that synaptonemal complex axial elements govern recombination pathway choice in mice.” (2011) (Li, Bolcun-Filas and Schimenti, 2011) 33
Figure 6 Expression of SYCP2 and HPV oncoprotein E6 of RNA samples from both HPV+ and HPV- cancer cell lines. Data is obtained from previous Dr. Tim Fenton, Dr. Peter Ellis collaboration student Cindy Dong. 35
Figure 7 Genome map of TOPO plasmid expressing full length SYCP2 43
Figure 8 SYCP2 expression across normal tissue. 50
Figure 9 CBioPortal(CBioPortal, 2013) plot showing SYCP2 expression across different cancers. 51
Figure 10 CBioPortal plot types of alterations of target gene SYCP2. 52
Figure 11 Western blot showing different protein lysates of HPV+ (SiHa) and HPV- (C33a) and mouse testis lysate as a positive control. The predicted size of SYCP2 is ~200 kDa. SYCP2 Ab. Conc. 1:1000 overnight incubation. Block 5% skimmed milk in TBST 0.1% Tween. 53
Figure 12 Western blot showing different protein lysates of HPV+ (SiHa) and HPV- (C33a) treated with MG132 in a time dependent manner [0hr, 30 min, 2hr, 3hr, 4hr] and mouse testis lysate as a positive control. The predicted size of SYCP2 is ~200 kDa. SYCP2 Ab. Conc. 1:1000 overnight incubation. Block 5% skimmed milk in TBST 0.1% Tween. 54
Figure 13 Western blot showing different protein lysates of HPV+ (SiHa) and HPV- (C33a) treated with MG132 in a time dependent manner [0hr, 30 min, 2hr, 3hr, 4hr] and mouse testis lysate as a positive control. The predicted size of SYCP2 is 200 kDa. SYCP2 Ab. Conc. 1:1000 2hr incubation. Block 5% skimmed milk in TBST 0.1% Tween. Washing steps done also in TBST 0.1% Tween. Exposure 49 seconds. 55
Figure 14 Western blot showing different protein lysates of HPV+ (SiHa) and HPV- (C33a) treated with MG132 in a time dependent manner [0hr, 30 min, 2hr, 3hr, 4hr] and mouse testis lysate as a positive control. The predicted size of SYCP2 is 200 kDa. (SYCP2 Ab. Conc. 1:1000) 2hr incubation. Block 5% skimmed milk in TBST 0.025% triton. Washing steps done also in TBST 0.025% triton. Exposure 3 minutes. 56
Figure 15 Western blot showing the MPER-insoluble pellets of HPV+ (SiHa) and HPV- (C33a) treated with MG132 in a time dependent manner [0hr, 30 min, 2hr, 3hr, 4hr] and mouse testis lysate as a positive control. The predicted size of SYCP2 is 200 kDa. Antibody concentration was SYCP2 1:1000 and overnight incubation. The blot shows

uncharacterised bands and none of the predicted molecular weight, confirming that the protein is not left behind with insoluble pellets during protein extraction. 57
Figure 16 Mouse testis microtubules with anti-SYCP2 antibody. Left: Positive control SYCP2 concentration of 1:400. Right: Negative control with only secondary antibody. This experiment was conducted to confirm that the purchased antibody was recognising the correct epitope of the gene of interest. To confirm the western blot results observed in (figure14) and (figure15), and eliminate the possibility that the bands observed might be uncharacterised bands. 58
Figure 17 KAPA HIFI PCR product. Starting product was 10ng of TOPO plasmid 60
Figure 18 pEGFP-N1 (third and fourth wells) after restriction digestion with Promega; BamHI and SalI and SYCP2 PCR product (last and second last wells) after restriction digestion with Promega; BamHI and SalI. 60
Figure 19 pEGFP-N1 plasmid expressing full length SYCP2 61

# List of Tables:

Table 1 Mammalian Cell-lines 36
Table 2 Primer pair designed for PCR cloning 41
Table 3 Cycle parameters for KAPA HiFi HotStart 42

# Abstract

High risk Human papillomavirus infections are linked to highly prevalent cancers; a consequence of aborted viral cycles and deregulation of HPV oncogenes E6 and E7. In many cases of cancer, meiotic genes can be seen upregulated, activation of meiotic genes in normal tissue can have detrimental effects due to the unique characteristics of gametogenesis that if observed in normal cells can be oncogenic. The gene of interest investigated in this project is SYCP2; it is a meiosis specific gene that is a part of the synaptonemal complex that links homologous chromosomes and facilitates synapsis. Recent studies showed it is strongly upregulated in HPV associated cancers when it's normal expression should be restricted to testis, it has been linked to what is called Cancer Testis Antigens (CTA) which are a group of genes that has been associated to tumours of different histological origins whose normal expression is restricted to testes. This gene is of interest because CTAs hold unique therapeutic potential as biomarkers or immunotherapeutics. It has been documented that SYCP2 is upregulated, however, existing data thus far only shows the mRNA level. In order to explore any therapeutic potential of this gene its expression profile needs to be investigated, and its possible correlation to viral infection or oncogenesis. In this project we aim to investigate if the large increase in SYCP2 transcript levels in HPV16+ cancers is paralleled by a similar increase in SYCP2 protein levels. This study has looked into the protein expression of SYCP2 that if further investigated, could form a useful model in understanding HPV infection and the possible role SYCP2 has in aiding oncogenesis.

# 1 Introduction

According to the World Health Organization (WHO) 2018 registered 570 000 new cervical cancer cases in the world, which accounts for 7.5% of female cancer mortality. Human Papillomavirus (HPV) is a DNA virus of the reproductive tract, HPV infection is attributed to sexual activity and found highest in females aged 30 to 34. There are over 100 strains of HPV with characteristic tropism, And around 12 types of HPV strains are considered high risk that lead to cancer. The HPV infection can be asymptomatic or can present as warts that are either Cutaneotropic in which they present as benign or malignant cutaneous and plantar warts, or Mucosotropic which present as benign or malignant genital warts. (Castellsagué, 2008). The viral infection can take up to 15 years to develop to precancerous lesions, preventative measures are advised; primary prevention by vaccination which is the best preventative approach to reduce global cervical cancer burden, and secondary prevention by screening using molecular biology techniques to detect the presence of the virus, the persistence and viral load per unit as well, as the type of HPV plays a role in the risk of oncogenesis. (*Human papillomavirus (HPV) and cervical cancer*, no date)

Scientific research over the last decades has demonstrated the causal effect of persistent HPV infection and cervical cancer, and has been claimed to be the first necessary cause of a human cancer ever identified. (Bosch *et al.*, 2002) (Castellsagué, 2008) HPV+ cancers present a unique molecular profile in regards to mutation burden, intratumoral heterogeneity as well as epidemiological profile. (Lechner and Fenton, 2016) Due to the

unique mutational signature of HPV+, in which the HPV oncoproteins introduce uniform mutations to host genes in the DNA repair machinery, this provides more targeted biomarkers and treatments which in turn provides more radiosensitivity. HPV+ cancers proved to have better clinical prognosis and better disease-free survival. Thus a patient's HPV status presents an advantage to the patient's response to treatment and survival. (Faraji *et al.*, no date) (Gillison *et al.*, 2019)

HPV presents a good model for studying alterations that lead to oncogenesis for the uniform and systemic mutations it introduces to host cells. This project focuses on a meiosis specific gene that has been documented to be upregulated in HPV+ cancers. We aim to investigate its possible role in HPV infection and/or tumour progression.

# Literature Review

## 1.1 Cancer Biology

Cancer is a collection of diseases that can affect any organ or type of cell. It is more a pattern in which cells deviate from the complex highly intricate structure of cellular maintenance. Essential Cell Biology Alberts B Hopkin K Johnson A et al. identifies cancer with two characteristics: (1) Proliferation in defiance of normal cellular constraints and (2) invading and colonising other cells' territories. (Alberts *et al.*, 2019)

Epidemiological studies have shown that environment and life style is an important role in the causation and progression of human cancers. And that shows in how different cancer types are more prevalent in different countries. For example, cervical cancer is more common in sexually active women and almost 99% of cervical cancers are caused by sexually transmitted virus HPV. (Moody and Laimins, 2010) However, even with the strong correlation of life style and environment, there is still much not known of the specific factors that drives oncogenesis, that could be due to the fact that studying cancer cells in vitro is not a true representation of the in vivo model. Especially the interactions and scaffolding of the extracellular matrix (ECM), that and the biochemical nature of the tumour microenvironment (TME) in different organs, such as pH, oxygen, nutrient and waste diffusion. The varied nature of TME and metabolic differences in each location or tissue adds the complexity of the physiology and subsequent severity of the cancer in a specific location, phenotype advancement and surrounding tumour stroma transformation

(Alberts *et al.*, 2019) (Burgos-Panadero *et al.*, 2019). For a cell to become cancerous it accumulates mutations; which are changes in the DNA sequence. This can happen through external stressors such as smoking, UV rays, radioactivity, and chemical carcinogens. However, even in an environment without these mutagens, mutations can still arise spontaneously through mistakes during the DNA replication process in cell division. Although DNA repair is a highly accurate and precise mechanisms mistakes may happen in which they introduce a mutation and this mutation would be carried out through this cells progeny (Alberts *et al.*, 2019).

### 1.1.1 Cancer and Epigenetics

The regulations and patterns of gene expression that do not entail any change in the DNA sequence and can be passed along to a cell's progeny is called epigenetic inheritance. A mechanism in which DNA expression is altered but without the change of the sequence is DNA Methylation in which a methyl group is added to the fifth carbon of cytosine residue to form 5-methylcytosine. This process is carried out by a family of DNA methyltransferase (DNMT), in addition to carrying out the formation of 5-methylcytosine they also during DNA replication carry out the DNA methylation pattern from parental DNA to daughter DNA. The specific process in which the DNMTs target these sites are unclear but there is a proposed mechanism in which RNA interference (RNAi) targets DNA sequences and silences them.

(Reece *et al.*, no date) This conserved process is what allows the variety of different gene expression in different cell types given the fact that essentially all cells share the same genome. When epigenetically silenced genes are activated there can be detrimental effects that lead to cancer. (Moore, Le and Fan, 2013)

Not all mutations cause cancer, there needs to be a very specific favourable environment for cancer cells to replicate as they do, such as the disruption of tumour suppressor genes, cell cycle regulating genes and promoting genetic instability. And it is with these characteristics, these cells would proliferate when they should not, and not follow the specific cellular instructions. It is not only one single mutation that would lead to tumour development it can take up to ten mutations. It is of importance to study the molecular environment in which these cellular violations happen to offer better more effective therapeutic options (Alberts *et al.*, 2019).

With the advancement of sequencing we have come to know the different mutational signatures of different tumours. Because cancer is not only one disease, analysing mutational profiles helps identifying driver mutations, and characterising the biology that creates the advantageous processes that ultimately promote tumorigenesis and help the discovery of biomarkers and personalised treatments.

## 1.2 Cancer Testis Antigens

Cancer Testis Antigens (CTA) are tumour associated antigens which are expressed in tumours of various histological origins. In normal tissue, CTA are almost exclusively expressed in the seminiferous tubules of the testes during spermatogenesis modulating meiosis. Epigenetic regulation might be the mechanism in which CTA expression is regulated, DNA methylation in particular. Epigenetic regulation protects cells from entering what is called 'soma-to-germline transition' which is when somatic cells gain cellular functions that are specific to germ line cells. When somatic cells gain meiosis specific functions, such as using interhomolog recombination, this gives rise to aberrant proliferation which in turn could lead to loss of tumour suppressor genes and genomic instability, which as previously described are hallmarks of cancer. CTAs have become a pharmacological target because of the nature of their tumour restricted pattern of expression that can allow for tumour directed immunotherapeutics. (Fratta *et al.*, 2011), (McFarlane and Wakeman, 2017)

### 1.2.1 Expression pattern and therapeutic potential of cancer testis antigens

In the early 1960s, the use of serological analysis and experimentations on mice showed that transplanting mice with certain tumours triggers an antigenic immune response. This finding cultivated an interest in the field because it showed the potential of employing the immune system's cytolytic activity to treat cancer (Whitehurst, 2014).

One of the first identified CTA family of antigens was by Van Der Bruggen and colleagues in 1991 termed MAGE-1. They are expressed in melanoma cancer cells, and further studies into this family of antigens showed the normal expression of this family of genes MAGE-1 is restricted to testis. (Van Der Bruggen *et al.*, 1991) These antigens are characterised by their mRNA expression being specific to germ cells, and their abnormal gene activation in non-germ cells is often linked with tumorigenesis. (Gibbs and Whitehurst, 2018) However, some have reported low expression in somatic tissue (<1% of their mRNA expression in testis) (Fratta *et al.*, 2011) There is reported low expression in the brain and reported presence in placenta and ovary and earlier stages of embryogenesis. (Whitehurst, 2014) (Costa, Le Blanc and Brodin, 2007)

Following the discovery of the first family of Cancer Testis Antigens MAGE-1 and their potential in immunotherapy, many other CTA family of antigens were discovered. The availability of cloning methods and use of cDNA libraries led to the identification of over 225 genes such as: GAGE-1 (Van den Eynde *et al.*, 1995), BAGE (Boël *et al.*, 1995) HOM-MEL-40 (Sahin *et al.*, 1995) SSX2 (Kawai *et al.*, 1998) and NY-ESO-1 (Chen *et al.*, 1997). The data showed that CTAs are more likely to be activated in lung, ovarian, bladder, melanoma and breast tumours, and less likely to be activated in leukaemia, lymphoma, renal, colon, and pancreatic tumours (Whitehurst, 2014).

In somatic tissue CTA are usually methylated, As mentioned previously; methylation is a molecular phenomenon where genes are silenced by the addition of methyl groups to the 5’

of cytosine in the cytosine-guanine dinucleotides in gene promoters. Catalysed by DNA methyltransferases (DNMTs), this process represses the expression of genes, with the help of chromatin remodelling co-repressor complex, binding methyl-CpG-binding protein (MBDs) (Figure 1). And in spermatogenesis CTAs are activated by demethylation. De novo expression of MAGE-A1 one of the first CTAs to be investigated correlated this pattern of expression in melanoma cell lines. The hypomethylation pattern of expression is also observed in different MAGE-A and NY-ESO when investigated in cell lines. This proposes an epigenetic alteration module of expression for CTA (Fratta *et al.*, 2011).

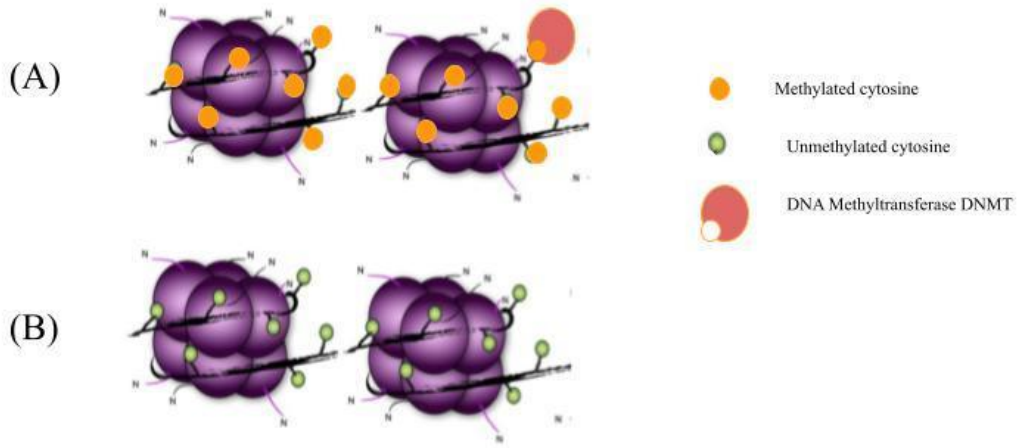


*Figure 1 Epigenetic regulation of CTA expression.*

*(A) CTAs are silenced in normal tissue, the chromatins are methylated and that is facilitated by DNA methyltransferase DNMT, binding of transcription factors is inhibited as there is no access to the promoter region. (B) In the absence of DNMTs transcription factors have access to promoter regions and result in transcriptionally active CTA genes. Figure adapted from:* (Fratta *et al.*, 2011)

CTAs are potentially a good target for immunotherapy as they showed an immunogenic response in recruiting the Cytotoxic T Lymphocytes (CTLs) in experimental models. Considering the low immunogenicity and immunosuppressive microenvironment of tumours makes for a good target for cancer vaccine. A study conducted in 2005 showed that infecting melanoma cells with ALVAC, a virus encoding MAGE antigens recognized by T-cells has showed that with repeated vaccination tumour regression and a CTL response is observed in a minority of melanoma patients (Van Baren *et al.*, 2005) (von Witzleben *et al.*, 2020).

Any gene whose expression is restricted to testis + neoplastic cells can be classified CTA. However, classification by expression data alone may not translate to the antigen's immunogenic status and its ability to elicit humoral and cellular immune responses and need to be further investigated. Expression patterns of testis-tumour genes and their possible therapeutic potential warrant further investigation, and in this project one specific testis gene will be investigated, SYCP2.

## 1.3 Human Papillomavirus

### 1.3.1 HPV structure and genome organisation

HPV is a circular double stranded non enveloped virus that is around 8000 bp, 50–60 nm in diameter and belongs to the *Papillomaviridae* family and weighs at 5x106 Dalton, it is evolutionary conserved with a divergence rate of 1% per 40,000-80,000 years. There are more than 240 types of HPV, and can be classified into low risk and high risk. 12 types are known to be associated with malignant lesions and drivers of oncogenesis. Of those types HPV-16, HPV-18, HPV-31 and HPV-33 are the most high risk types, making up 75% of all HPV associated squamous cell carcinoma and 94% of adenocarcinomas. 30% of females exposed to oncogenic HPVs are infected within 24 months post sexual exposure, 90% clear within 2 years but persistent prolonged infection may lead to cervical intraepithelial neoplasia (CIN) (Araldi *et al.*, 2018a) (Serrano *et al.*, 2015) (McBride, 2017c)

The virus has 8 Open Reading Frames (ORF) which through the method of alternative splicing encode the following genes; E1, E2, E5, E6, E7 and capsid proteins L1 and L2. These proteins help make an environment where a productive viral life cycle can be achieved. (Van Doorslaer, 2013) (Araldi *et al.*, 2018a) The genome as shown in (Figure 2) contains three regions; Upstream Regulatory Region (URR) that controls transcription and early region encoding the early proteins and lastly the late region encoding the late proteins.(McBride, 2017c)

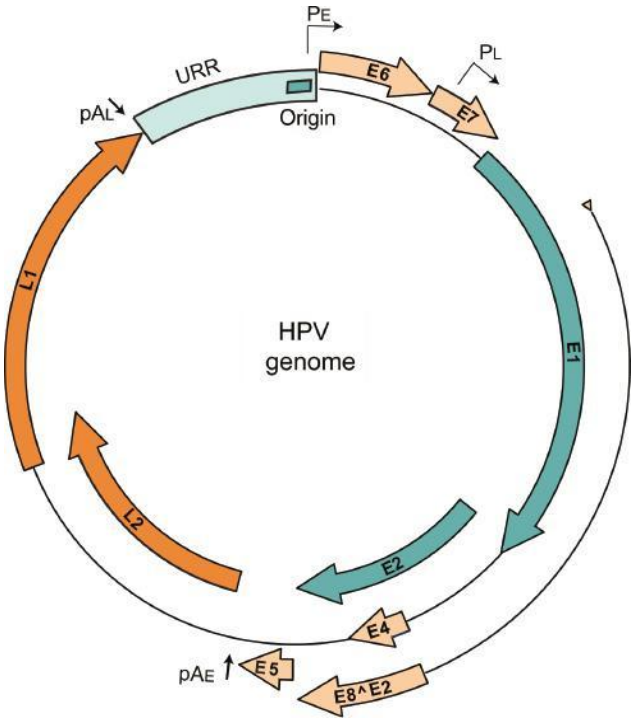


*Figure 2 Figure shown is the genome of circular double stranded DNA of Human Papillomavirus. Photo Credit: McBride, Alison A. "Mechanisms and strategies of papillomavirus replication." Biological chemistry vol. 398,8 (2017): 919-927. doi:10.1515/hsz-2017-0113* (McBride, 2017a)

### 1.3.2 The HPV life cycle

i. Viral Entry

The method of infection of HPV is through the epithelia, skin cells, oral and genital mucosa. Sexual contact is its known method of infection and so an individual's sexual behaviour and possible multiple sexual partners can be attributed to be the risk factor. However in the case of HPV-42 and HPV-16 there has been reported vertical mode of transmission; from mother to new-born. The virus gains entry through micro abrasions and infects the basal cells, taking advantage of the wound healing process which is an environment of high proliferation and self-renewing cells. The capsid protein of the virus

L1 binds to heparin sulphate proteoglycans on the basement membrane inducing a conformational change allowing the virus to bind to the surface of the basal cells. (Day and Schelhaas, 2014)

Through the process of endocytosis the virus is engulfed, where the capsid protein L1 is removed and only the HPV mini chromosome L2 complex remains. And through endosomal trafficking the HPV genome-L2 complex finds its way to the trans-Golgi network. When the cell undergoes mitosis and the nuclear envelope breaks down the L2-HPV genome complex gains access to the nucleus.

ii. Viral genome replication in the basal layer

HPV generates most of its progeny virions in stratified epithelia, however, it starts its virus life cycle in the basal cells generating in low copy numbers and as the cell differentiates and moves up the epithelium otherwise known as differentiation dependent life cycle, it takes advantage of that process and proliferates more copies of the virus and packages them, and when the cells are shed as a part of the natural tissue renewal process the newly packaged virions are released and find a new host. (Figure 3) (Stanley, 2012) (Chow, Broker and Steinberg, 2010)

As mentioned previously the HPV genome-L2 complex gains access to the nucleus after mitosis and nuclear envelope breakdown, during that process the HPV genome-L2 complex is observed localising alongside Nuclear Domain 10 (ND10) nuclear bodies and to create a

microenvironment favourable for viral transcription and replication, L2 changes Daxx and sp100 localisation in ND10 body to allow for transcription and replication of HPV DNA. (McBride, 2017b) In this first stage viral infection 'establishment' in the basal layer the virus and specifically viral genes E1 and E2 mediates constant low level of transcription and DNA replication as an extrachromosomal circular genome and it is passed down to daughter cells with every cell division. This conserved process along with the tethering of the HPV genome-L2 complex to immune hotspots such as the ND10 nuclear body is thought to be why HPV can escape immune recognition. (McBride, 2017c)

Early region genes encode the early proteins. E1, E2, and E4 HPV proteins, which as mentioned previously are essential to initiate virus replication. With E1 being the most conserved sequence in the whole of the papillomaviruses, E1 interacts with E2 to form a complex called the E1-E2 complex and initiates ori dependent replication. E1 also acts as a helicase binds and unwinds DNA allowing for access to the replication process. E1 initiates viral replication when its N-terminal domain binds to the motif region of Cdk2(Ma *et al.*, 1999). The complex initiated by E1-E2 attracts topoisomerase 1, and DNA polymerase α and replication protein A (RPA) (Araldi *et al.*, 2018a). An important role for E2 is that it acts as a transcriptional regulator for HPV oncoproteins E6 and E7 and regulates the amount of virions produced in a cell as it is responsible for the recruitment of the transcription factors required to initiate replication ensuring that only limited number of viral copies are being amplified during the early stages (Cai *et al.*, 2013).

iii. Viral genome maintenance

The viral life cycle of HPV goes through initial amplification to establish the steady low copy replication maintained in the second stage; maintenance. Which is carried out in constant low copy numbers as an extrachromosomal episome providing indefinitely a pool of viral infected cells to ensure persistence infection. They are estimated to be around 20 viral copies per cell.(McBride, 2008)

Furthermore, HPV protein E2 plays a role in the passing down of viral DNA to daughter cells by forming a complex with the protein Brd4. E2 forms the complex by binding the Brd4 protein to its C-terminal domain forming a complex that tethers the viral genome to the host chromatin, this process is what allows persistent infection by making sure of successful partitioning of DNA through the stable association with host chromatin in which they would remain in the nucleus membrane after cell division and pass down the HPV genome to daughter cells. (Helfer, Yan and You, 2014)

An important result of the interaction of Brd4 and E2 is binding to common fragile sites and because of the unique property of these sites E2 is able to induce, hijack and facilitate the DNA damage response (DDR) to produce the viral progeny (Jang, Shen and McBride, 2014)

iv. Viral Amplification and Packaging

As the cell starts to differentiate and move up the epithelium, the virus takes advantage of the differentiation process and employs the host cell machinery to produce its viral progeny. Helping this process is spliced mRNA product viral protein E4, E1^E4 it interferes with the cell cycle to achieve the S phase like state to allow for viral DNA amplification (McBride, 2008). HPV oncoprotein E7 drives the host cell into S phase and E1^E4 arrests the cells in G2 phase as it is where the virus can achieve higher viral DNA amplification. In this stage the DDR pathway is activated by an E7 induced interaction with ATM, and the DNA repair machinery in the DDR pathway is hijacked and used to amplify viral DNA in a favourable site via the interaction of E2-Brd4 mentioned previously (Moody and Laimins, 2009).

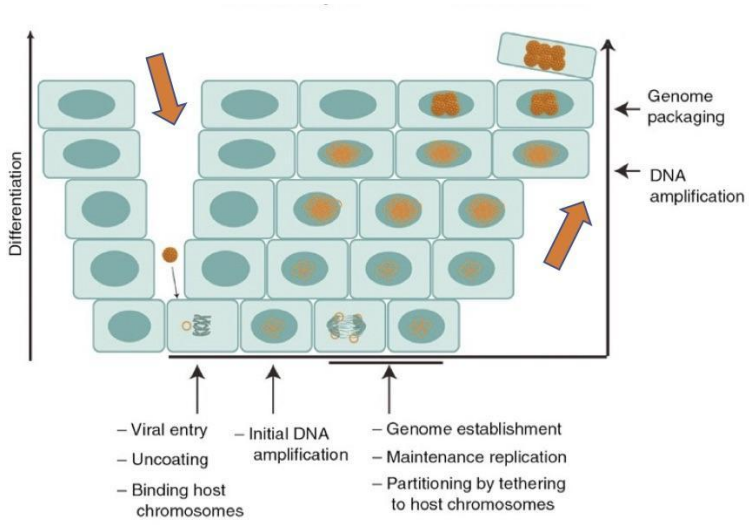


*Figure 3 The different stages of the papillomavirus life cycle.*

*The figure shows cells of differentiated stratified epithelium, the arrow pointing downwards shows how HPV gains access to the cells of the basal layers through micro abrasions. Which is where HPV begins initial DNA amplification and establishes maintained viral genome replication in sync with host genome replication. The arrow pointing upwards shows as the infected cells begin to differentiate, they begin amplifying viral DNA at a higher number and begin viral DNA packaging and the assembly of the virion capsid, and thereafter viral laden cells are sloughed off ready for a new host. Photo Credit: McBride, Alison A. "Mechanisms and strategies of papillomavirus replication." Biological chemistry vol. 398,8 (2017): 919-927. doi:10.1515/hsz-2017-0113*
(McBride, 2017a)

As mentioned, Viral DNA amplification happens in G2 arrested cells, and not in S phase where the host DNA is amplified. The reasons behind the viral DNA amplification happening after host DNA amplification are not widely understood, but it has been suggested that the host cell replication machinery is easier to hijack and can amplify DNA efficiently and more rapidly when it is not engaged with the replication of the host DNA (Wang *et al.*, 2009). Highest viral amplification occurs in terminally differentiated cells, the reason behind that could be that differentiated cells have less immune surveillance (McBride, 2017a). And subsequent to that the morphogenesis of progeny virions is observed.

The viral capsid packaging is facilitated by the late region genes, which encode L1,L2 the major capsid proteins. L1 protein mediates the binding of the virion to heparin sulphate receptors that are on the surface of the basal membrane giving the structural characteristics that allows it to invade cells. L2 is important in the first stages of the viral cycle as well as the last, as it mediates the DNA packaging and the assembly of the virion capsid. (Buck, Day and Trus, 2013) (McBride, 2017a) In the final stages in terminally differentiated cells, the packaged virion progeny is found throughout the nuclei, the viral genome is condensed into chromatins with host histones inside the viral capsid. And with the tissue renewal process these viral laden cells are sloughed off ready to infect a new host.

### 1.3.3 HPV induced Oncogenesis

Most HPV present as benign lesions or (papilloma), but some high risk HPVs have evolved through persistent infection and interfering with cellular pathways leading to onset of oncogenesis. Which include but not limited to cervical cancer, head and neck cancer and oropharyngeal squamous cell carcinoma (OPSCC). As previously discussed HPV genetic amplification happens with alteration in the cell cycle of through recruited DDR, in that when the differentiated cell should have already exited the cell cycle, for that proliferative state to be maintained E7 disrupt many important cell cycle regulators disrupting pathways such as pRb, p107, p130. And a consequence of HPV oncoprotein E7 through these molecular events leads to the activation of p53 tumour suppressor pathway.

E6 HPV oncoprotein plays a role in cell immortalization, by causing an upregulation in the expression of the human telomerase reverse transcriptase (hTERT) gene. And it also upregulates the expression of the NKX2-1 which in turn upregulates FOXM1 in HPV associated head and neck carcinoma that induces B-catenin nuclear translocation, these series of events often lead to the proliferation of keratinocytes and subsequent tumorigenesis (Araldi *et al.*, 2018b). Furthermore, E6 similar to E7, would too alter the effects of the tumour suppressor protein p53. This happens through E6 forming a ternary complex with E6AP ubiquitin ligase and p53 leading to the degradation of p53.

Mutations in p53 are linked to uncontrolled proliferation, and developmental defects. This phenomenon is linked to tumour progression because of the fundamental role p53 plays in

regulating gene expression, p53 responds to acute stress cell signals which are often a sign of oncogenic progression by driving the cell into programmed cell death. (KH and DP, 2007),(Rivlin *et al.*, 2011).

In addition, E6 can play a role in the genetic instability of the host genome by interacting with proteins responsible for single strand DNA break repair, and by that introducing mutations leading to cancer initiation. E6 can work with E7 to induce mutations and genetic instability as their co-expression has been shown to induce extracellular matrix remodelling, they also work to evade the immune system by downregulating Toll Like Receptors (TLR), and thus escape virus recognition and the activation of phagocytes (Araldi *et al.*, 2017) (Cai *et al.*, 2013).

The E5 gene encoding E5 HPV oncoprotein is a hydrophobic transmembrane protein; this protein is associated with cancer progression because of its mitogenic activity. The oncoprotein E5 can trigger cellular proliferation by changing the endosomal ph from 5.9 to 6.9 which inhibits the degradation of the Epidermal Growth Factor (EGF) and it can work together with EGF to recycle back the receptors from endocytosis to the cell membrane. E5 loses its Open Reading Frame (ORF) after genome integration which further suggests that it is involved in the early processes of cancer progression.(Gutierrez-Xicotencatl *et al.*, 2020) E5 can also be found on the golgi apparatus (GA) and Endoplasmic Reticulum (ER). E5's interaction with GA interrupts the major histocompatibility class I encoded protein human leukocyte antigen (HLA) which are important for immune regulation, interrupts its

expression on cell surface, this allows the virus to evade immune recognition and thus allow for viral-infection persistence. In addition, E5 uses Bax ubiquitination to inhibit apoptosis in cervical cancer; this leads to oncogenesis (Araldi *et al.*, 2018b).

This continued viral persistence and maintenance of the transformed state of the host cell when it should have exited the cell cycle, with the aforementioned characteristics of genetic instability could lead to genomic integration. E1 also aids with HPV viral integration by eliciting DNA breaks in the host genome. And this result is not necessarily a part of the virus life cycle because viral genomic integration might actually be considered the dead-end of a productive viral infection as no progeny virions can be produced going forward.

As mentioned previously E2 acts as a transcriptional regulator for E6 and E7, facilitating cell immortalisation and evading immune recognition. When the HPV genome is integrated onto the host genome E2 regulation is disrupted and a dysregulation in the expression of E6 and E7 HPV oncogenes is observed, and what it creates of an environment of uncontrolled proliferation, and genetic instability and these are often the hallmarks of cancer progression (Jang, Shen and McBride, 2014) (McBride and Warburton, 2017).

HPV disrupts a number of host genes, investigating these alterations opens a myriad of possibilities such as vaccines, biomarkers for early detection, and personalised treatments through neoantigens; where cancer treatment allows for the eradication of cancer cells with

less harm to healthy cells. In this project we investigate a meiosis specific gene upregulated by HPV, *SYCP2*.

## 1.4 Synaptonemal Complex

Meiosis is the molecular process observed in sexually productive organisms resulting in spermatozoa and oocytes in the case of mammals. What distinguishes this process from mitosis is that instead of the production of genetically identical diploid daughter cells, meiosis results in haploid gametes, with one copy of either paternal or maternal chromosomes. The unique and defining event in meiosis is the highly conserved and controlled molecular process called crossing over, where maternal and paternal chromosomes are tethered by a protein complex called the Synaptonemal Complex pair in an X shaped structure called the chiasma facilitating meiotic recombination where programmed DSB are initiated and repaired. This genetic exchange observed in meiosis is allowed to happen because of the favoring of inter-homologous DNA repair as opposed to inter-sister DNA repair observed in mitosis.

### 1.4.1 The Synaptonemal Complex role in DNA replication

During the cell cycle, the cell undergoes double strand breaks (DSB) from different stressors; ionizing radiation, reactive oxygen species, Ultraviolet light, DNA replication errors, as well as regulated DSB to facilitate recombination in meiosis induced by SPO11, in which DNA DSB repair pathways are recruited. DNA repair pathways include; Non homologous End-Joining (NHEJ) which is predominantly used during cell cycle, it does not need a homologous template and is facilitated by directly re-ligating the broken ends

together. If NHEJ fails and there is damage from the broken ends, Microhomology Mediated End Joining (MMEJ) is recruited which includes resection of the broken ends to mediate homology. This pathway is associated with deletions and chromosomal abnormalities. (Pardo, Gómez-González and Aguilera, 2009) Homologous Recombination (HR) DNA repair pathway, which includes inter-sister and inter-homologue recombination, relies on the access of a homologous template. In mitosis we witness a favouring of inter-sister HR, it is present in the G2 phase of the cell cycle because that is when the chromosomes are duplicated. In meiosis I, we witness inter-homolog HR bias, this occurs in the pachytene stage of prophase 1, inter-homolog recombination is important in meiosis because it creates genetic variety to the offspring. The Recombinase proteins facilitate recombination (Rad51-Rad54 in eukaryotes) by searching for homology and then, promoted by HORMAD-1 (meiotic HORMAD proteins regulate DSB and crossover formation) ‘presynaptic alignment’ is then formed and the synaptonemal complex (SC) assembles (Figure4) (Pardo, Gómez-González and Aguilera, 2009).

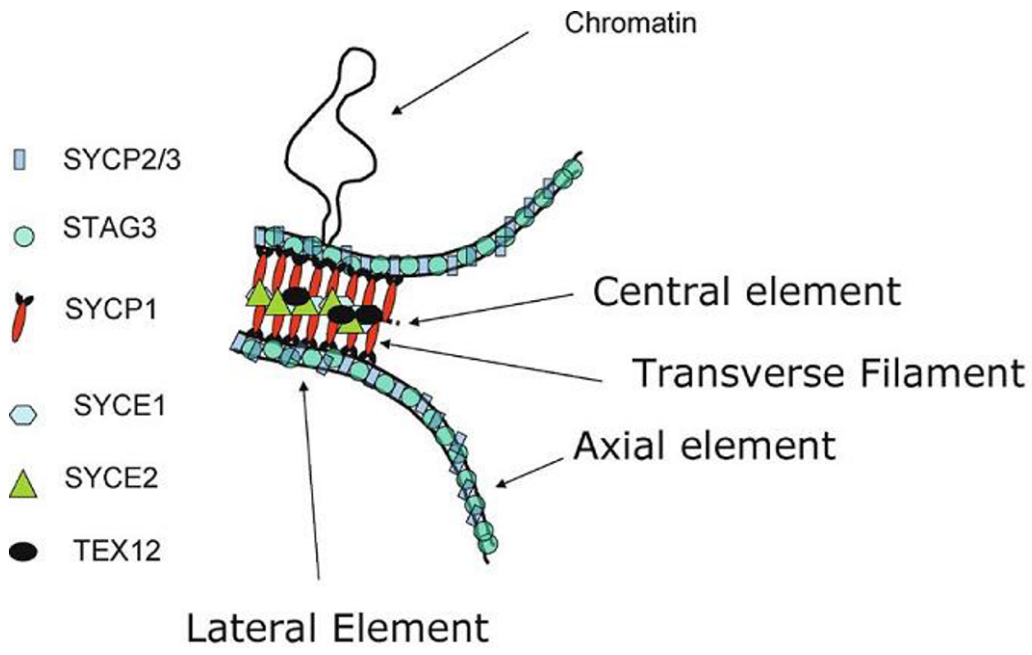


*Figure 4 Model of the elements of the synaptonemal complex. Photo Credit: Molecular Biology of the Cell by Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter, Peter New York and London: Garland Science; c2002*

### 1.4.2 The importance of the Synaptonemal Complex for chromosome alignment during meiosis

The Synaptonemal Complex comprises of; axial elements (AE) precursors of lateral elements (LE) in synapsis, and transverse filaments (TF) to make up the mature SC. Synapsis occurs by: AE forming between sister chromatids, the AEs of two homologous chromosomes are connected by the TFs and when SC disassembles, homologous chromosomes are separated except at crossover regions which are called chiasmata. (Del Val *et al.*, 2019) (nd Kelsey C. MartinMhatre V. Ho, Ji-Ann Lee, 2012) (West *et al.*, 2019) Components of this proteinaceous structure are: SYCP1; SYCP2 and SYCP3. Together

they form a heterotetrameric coiled-coil complex that self-assembles into extended filaments. Cohesin complexes interact with HORMAD proteins and SYCP2/SYCP3 and make up the mammalian meiotic chromosome axis. This process is conserved across fungi, mammals, and plants.

Mutations in SC are linked to male infertility, the AEs fail to form and no synapsis is witnessed and that leads to improper segregation of chromosomes at meiotic metaphase I, creating genomic instability and leads to aneuploid gametes.(Yang *et al.*, 2006) (Geisinger and Benavente, 2017)

SYCP3 is required for the recruitment of SYCP2 but not SYCP1, they are structural components of AE/LE. The synaptonemal complex facilitates synapsis because it is a part of the chromosome axis core proteins, and seen during the pachytene stage of meiosis I because it links homologous chromosomes together and aids crossover.

### 1.4.3 The Synaptonemal Complex's role in suppressing inter-sister recombination during meiosis

In addition, a study published in 2011(Li, Bolcun-Filas and Schimenti, 2011) explored that SCs role in meiosis is not only a structural role and mediating proper chromosome segregation, but that the SC also plays a role in recombination partner choice by creating

inter-homolog bias in the repair of programmed meiotic DSB (West *et al.*, 2019) (Yang *et al.*, 2006) (Li, Bolcun-Filas and Schimenti, 2011) .

The study started with the observation that a proportion of oocytes in SYCP3-deficient female mice nevertheless survive despite the fact that the SC is unable to assemble. This is counterintuitive, since in the absence of the SC normal meiotic recombination cannot occur. Thus, oocytes from SYCP3-deficient females would be expected to show persistent unrepaired DNA damage, triggering apoptosis.

Their first hypothesis was that the SC plays a role in the DNA damage checkpoint. In this case, this check point would be lost in SYCP3-null oocytes, allowing them to survive despite persistent DNA damage. To test that theory they created a SYCP3 mutant, a Rec8 mutant and a SYCP3, Rec8 double mutant. A Rec8 mutant inhibits the sister chromatids from binding and subsequently prevents synapsis between homologs which in turn prevents all recombination inter-sister or inter-homologous and subsequent repairs to DNA damage, therefore spermatocytes and oocytes would be eliminated by the DNA damage checkpoint. If the SC plays a role in the DNA damage checkpoint then SYCP3 deficiency should be able to rescue the effects of Rec8 mutation, allowing oocytes to survive in the double mutant females. However, the double mutant showed no viable oocytes indicating that these have a functional DNA damage checkpoint despite lacking SYCP3.

This led to an alternate hypothesis: that when SYCP3 is knocked out DSB are repaired by NHEJ instead of by HR. To test this second hypothesis, they used the severe combined immunodeficiency (scid) model which lacks the PRKDC gene required for the non-homologous end joining (NHEJ) pathway of DNA repair. If DSBs in SYCP3-deficient oocytes are repaired by NHEJ, then the SYCP3, scid double mutant should be unable to repair meiotic DSBs and this double mutation should prevent oocytes survival. However the SYCP3, scid double mutant showed viable oocytes and that eliminated the theory that NHEJ repairs DSB introduced by SPO11 in meiosis.

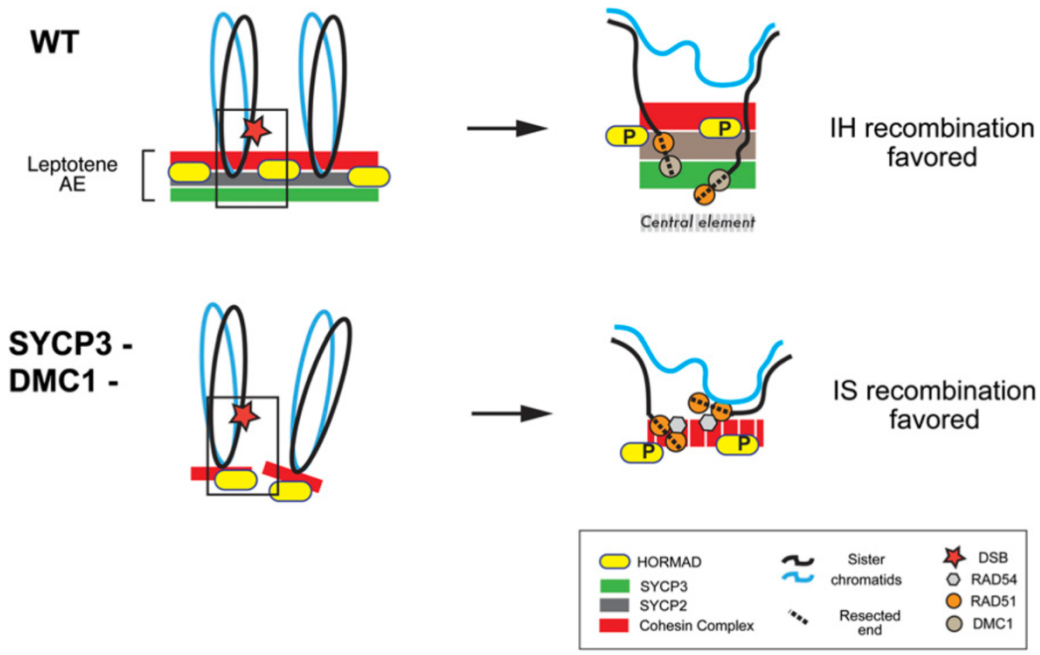


*Figure 5 Model for role of axial element in promoting interhomolog recombination bias proposed by Li, Xin Chenglin et al. "Genetic evidence that synaptonemal complex axial elements govern recombination pathway choice in mice." (2011)* (Li, Bolcun-Filas and Schimenti, 2011)

This led to their third and final hypothesis that SYCP3 affects the choice of template used for DNA repair by HR. To test the third hypothesis they used Rad54 and Dmc1 mutant and double mutant females in combination with SYCP3 mutations. RAD54 is essential for

inter-sister recombination, while Dmc1 is essential for inter-homolog recombination. Dmc1 mutant females lack oocytes, showing that the normal pathway for DNA repair in oocytes strictly requires inter-homolog HR. The DMC1/SYCP3 double mutant is viable, but with a lower oocyte count. This shows that SYCP3 must facilitate the repair of meiotic DSBs by some pathway other than inter-homolog HR. The DMC1/RAD54/SYCP3 triple mutant lacks oocytes, showing that the rescue of oocyte loss in the DMC1/SYCP3 double mutant must require a functional pathway for inter-sister HR (Figure 5).

The conclusion is that one function of the synaptonemal complex during meiosis is to suppress inter-sister HR, and that loss of the synaptonemal complex permits the repair of meiotic DSBs by inter-sister HR. Although focused on Sycp3, this study did also investigate a double mutation of Sycp2 and Trip13, and concluded that loss of either Sycp2 or Sycp3 can relieve the block to inter-sister HR.

### 1.4.4 The Synaptonemal Complex gene SYCP2 expression in cancer

The gene investigated in this project SYCP2, is of interest because of its upregulation in HPV associated cancers. Published data analysing RNA expression by conducting wide RNA sequencing for transcripts with possible statistical significance found that *SYCP2* showed the most significant upregulation from baseline to premalignant tissue; higher than *CDKN2A*, which is used as a clinical biomarker for HPV in OPSCC (Slebos *et al.*, 2006) (Masterson *et al.*, 2015).

Furthermore, our research group previously demonstrated that SYCP2 is part of a set of genes that are consistently upregulated in HPV+ cancers, irrespective of the anatomical location at which they arise (Chakravarthy *et al.*, 2016)

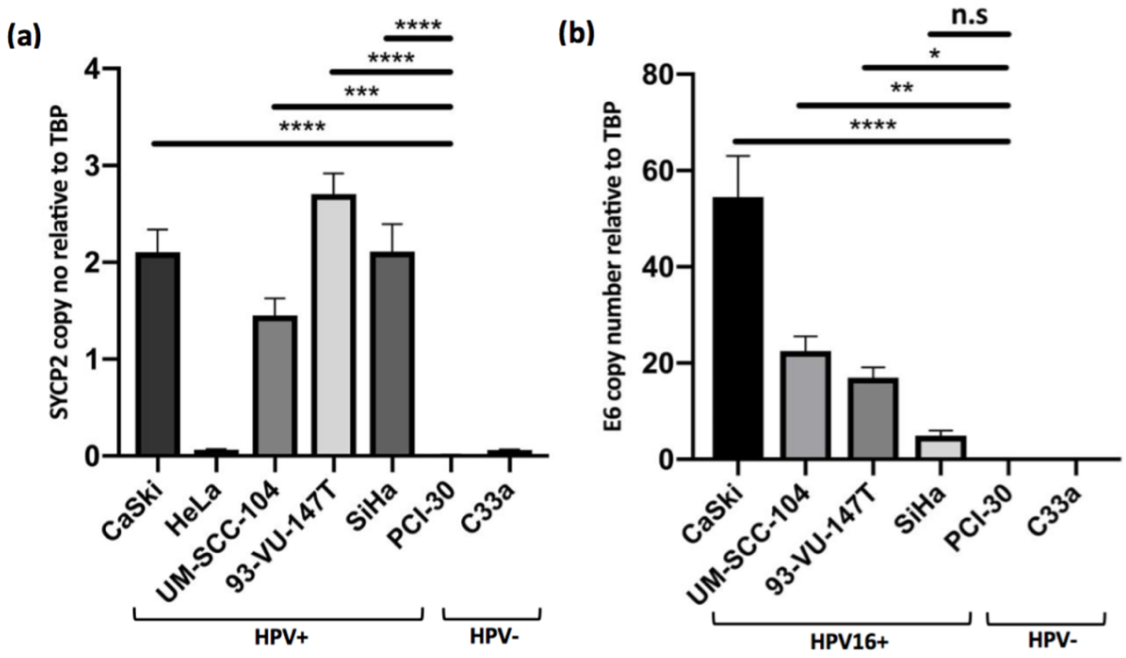


*Figure 6 Expression of SYCP2 and HPV oncoprotein E6 of RNA samples from both HPV+ and HPV- cancer cell lines. Data is obtained from previous Dr. Tim Fenton, Dr. Peter Ellis collaboration student Cindy Dong.*

Data from genomic analysis of cervical cancers caused by different HPV types by the Fenton lab (Chakravarthy *et al.*, 2020) suggests SYCP2 upregulation is specifically seen in cervical cancers caused by HPV types in the alpha-9 clade which include HPV16, but not in the alpha-7 clade, of which HPV18 is a member. Moreover, previous MSc-R student (Cindy Dong) confirmed using qRT-PCR, the upregulated expression of SYCP2 in HPV associated cancer cell lines (Figure 6). As can be seen in the figure there is significant

upregulation in HPV 16 cancer cell lines. However, existing data only shows expression on the mRNA level,

There is no existing data assessing the protein expression of SYCP2, confirming that mRNA expression is followed by protein expression would be a stepping stone into exploring the possible role this gene may have in viral infection or cancer progression or both. Which is why it is one of the aims of this project. Exploring the expression pattern of the gene interest can open the possibility of finding a non-invasive screening tool for HPV specific cancers. Moreover, existing literature does not provide sequenced analysis of final protein structure of the wrongfully expressed form of SYCP2, it could be that the protein product being expressed in HPV+ cancer is an alternatively spliced form of SYCP2, as opposed to the much larger full length expressed in testes. (Shakib *et al.*, 2005) (Guo *et al.*, 2020)

# 2 Aims

As presented through the literature review; our gene of interest has been linked to HPV+ cancers. The mRNA data available showing the significant upregulation from baseline to premalignant tissue warrants further investigation, considering the potential it has in creating a non-invasive screening tool. In view of this, we present the following hypotheses:

Hypothesis 1: Overexpression of SYCP2 on the mRNA level is linked to overexpression of SYCP2 on the protein level.

Hypothesis 2: Noncanonical expression of SYCP2 is linked to oncogenic progression.

We aim to test the presented hypotheses by exploring the available genomics data using bioinformatic tools. By conducting a bioinformatic review of SYCP2 expression across different tissues, illustrating the differential expression in normal and cancer tissue. Furthermore, Using a wet lab approach investigate the protein expression pattern of SYCP2 in HPV associated cancer cell lines, and on that explore if the noncanonical upregulation of SYCP2 mRNA in HPV+ cancer cell lines is linked to upregulation in SYCP2 protein. And lastly, to investigate the role of SYCP2 in sustaining HPV infection, by the use of SYCP2 knockout cells.

# 3 Methodology

## 2.1 Bioinformatics:

Online bioinformatics tools were accessed to explore the expression profile of the gene investigated in this project including: GTEx Portal(*GTEx Portal*, no date) which is a database encompassing different expression profiles of genes in normal tissue. And CBioPortal for cancer genomics (CBioPortal, 2013)

## 2.2 Tissue Culture:

### 2.2.1 Cell lines

The cancer cell lines in this study (see Table 1) were cultured with Dulbecco's Modified Eagle Medium (DMEM) supplied with 5% foetal bovine serum (FBS) and 1% penicillin/streptomycin. And incubated in 37℃ and 5% CO2. Media was changed according to the state of the cells 48-72 hours. And when the appropriate confluency of cells has been reached it is then passaged; washed with 1x Phosphate Buffered Saline (PBS) followed by 0.05% Trypsin-EDTA and when the cells have detached, diluted with DMEM and spun down 300 G for 6 minutes and resuspended in fresh DMEM and plated onto fresh T75 culture flasks.

| Cell line | Cell Origin | Cell Type | Morphology | HPV Status |
|---|---|---|---|---|
| C-33A ATCC® HTB-31™ | Homosapien - Cervix | Epithelial, Retinoblastoma | Adherent | HPV-Negative |
| SiHa ATCC® HTB-35™ | Homosapien -Cervix | Grade-II Squamous Cell Carcinoma | Adherent | HPV-Positive |

*Table 1 Mammalian Cell-lines*

## 2.3 Protein Analysis:

### 2.3.1 Protein Extraction

Cells were grown in 6-well plates 500,000 cells per well to be treated with MG132 which is a proteosome inhibitor to stabilise the protein and avoid degradation. The cells were treated in 6 different time intervals; 25uM of MG132 4 hours, 3 hours, 2 hours, 1 hour, 30 minute, and a control; 25uM DMSO. To extract the protein; first the cells were washed with cold 1x PBS on ice to ensure low protease activity and then the cells were lysed using Lysis buffer ( 960uL MPER + 40 uL Protease inhibitor + 0.4 uL Benzonase) 50 uL for each well, using a cell scraper the cellular contents collected in 1.5 mL Eppendorf tubes lysed on ice for 15 minutes. Following that, the samples were spun down in a 4℃ centrifuge for 20 minutes at

10,000 rpm. The supernatant collected and put in final 1.5 mL Eppendorf tubes ready for Protein Analysis. And pellets saved for later experiments.

### 2.3.2 Protein Quantification

Obtained protein samples were quantified for subsequent experiments, for this the ThermoFisher Pierce™ BCA Protein Assay Kit was used. This Assay helps measure protein concentrations in whole cell lysates to allow for calculation of final ug/uL concentrations.

### 2.3.3 Samples prep

Quantified samples were prepared for western blotting by mixing with 4x Laemmli buffer: [final conc. 277.8 mM Tris-HCL PH6.8 , 4.4% SDS , 44.4% glycerol , 0.02% Bromophenol blue] + 1M Dithiothreitol (DTT) samples mixed and placed in heat block for 10 minutes in 95℃.

### 2.3.4 SDS PAGE

The samples were loaded with 5 uL of (Thermo Scientific PageRuler™ Plus Prestained Protein Ladder) run in a 8% hand-cast gel of which final concentration is [40% Acrylamide

, 1.5M Tris-HCL PH 8.8 , 10% SDS , 3.75 uL TEMED , 10% APS] And run in Running buffer (Invitrogen NuPAGE™ MOPS SDS Running Buffer) at 150v for 1 hour.

2.3.5 Western Blotting

*Multiple approaches were attempted to optimise transfer.*

a. Semi-dry transfer

The samples were transferred on a PVDF membrane using the BioRad Trans-Blot Turbo System for 20 minutes following manufacturer protocol.

b. Wet transfer

The samples were transferred to PVDF membrane via a wet tank transfer using Towbin Transfer Buffer (Towbin et al. (1979) PNAS USA 76:4350-4) (Towbin, Staehelin and Gordon, 1979) is the buffer of choice for tank blotting; 25 mM Tris, 192 mM glycine, 10-20 % methanol] and left overnight on 40v in 4℃.

After blocking the PVDF membrane in 1x Tris-buffered saline with 0.1% Tween (TBS-T) and 5% skimmed milk for 1 hour on a rocker at room temperature it was followed with 2 hour anti-SYCP2 antibody incubation (1:1000 dilution in blocking buffer) on a rocker in

4℃ (Millipore Anti-SYCP2 - Polyclonal Antibody. Cat. #ABE2622) And after that it was washed 5x5 (5 times for 5 minutes) with TBS to prepare for horseradish peroxidase (HRP) conjugated secondary antibody (Donkey Anti-rabbit Invitrogen Cat. #A16023 , 1:10,000 in blocking buffer) in room temperature for 1 hour. And after that it was ready for protein detection using Syngene G-Box fluorescence imager, and for that a HRP substrate luminol is used (ECL) (Biorad #1705061)

## 2.4 Gene Cloning

### 2.4.1 Bacterial Culture + DNA Purification

A glycerol stock of plasmid pCR-XL-TOPO (Figure 7) expressing full length SYCP2 was purchased, (Cat. # ABIN3996587) Bacterial resistance: Kanamycin.

Aforementioned glycerol stock was grown on agar plate supplied with 50 ug/mL Kanamycin antibiotic, using sterile techniques. And left to grow in 37℃ incubator overnight. One colony was chosen and introduced in LB broth with 1:1000 dilution of Kanamycin antibiotic, and left in shaking 37℃ incubator. Finally the liquid culture is purified to isolate plasmids using Qiagen QIAprep spin Miniprep Kit (Cat. #27106). Purified Plasmids were then quantified using Nanodrop and was a 145ng/uL.

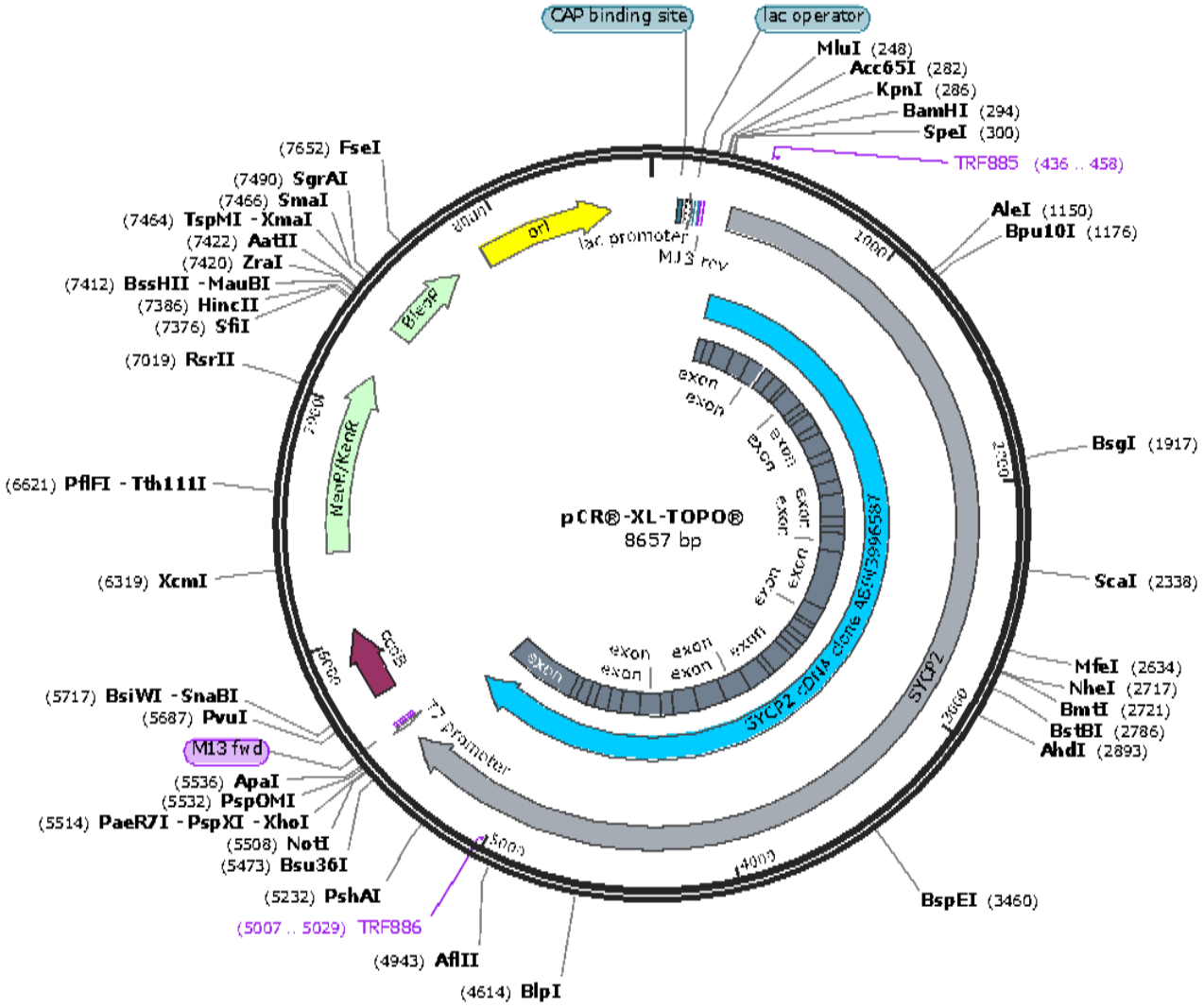


*Figure 7 Genome map of TOPO plasmid expressing full length SYCP2*

2.4.2 Designing Primers

Forward and Reverse primers were designed for subsequent gene cloning experiments.

(Table 2) Using Integrated DNA Technologies (IDT)

| Primer | Tm | Sequence | Size in bp |
|---|---|---|---|
| Forward 5’ to 3’ | 61.8℃ | 5’-GAT CGT CGA CCA AAA TGC CAA TAA GAC CAG ATC-3’ | 33 |
| Reverse 5’ to 3’ | 59.1℃ | 5’-GAT CGG ATC CAC ATT AGC ATT TCT TTC ATG AG-3’ | 32 |

*Table 2 Primer pair designed for PCR cloning*

2.4.3 Polymerase Chain Reaction PCR

The PCR reaction were done using KAPA HiFi PCR kit and conducted per manufacturer’s data sheet recommendation; cycle protocol is as follows (Table 3)

The reactions were done in triplicates and a water control, all on ice. Every reaction was supplied with 0.3 μM of each forward and reverse primers, and 5uL of KAPA HiFi HotStart ReadyMix.

The starting DNA sample was 10ng each. After the PCR round has finished to test whether the protocol has worked 1 uL of each reaction was run on gel electrophoresis 1% Agarose gel was chosen for better visualization of larger DNA [ 50 mL TAE + 0.5 g of Agarose + 2.5 uL Ethidium Bromide]. Samples were loaded with Invitrogen™ 1 Kb Plus DNA Ladder

(Cat. # 10787018) run for 1 hour at 80 volts. The bands were cut off and purified using New England BioLabs Monarch® DNA Gel Extraction Kit (Cat. #T1020)

| Step | Temperature | Duration | Cycles |
|---|---|---|---|
| Initial Denaturation | 94℃ | 3 minutes | 1 cycle |
| Denaturation | 98℃ | 20 seconds | 30 cycles |
| Annealing | 60℃ | 15 seconds | 30 cycles |
| Extension | 72℃ | 3 minutes | 30 cycles |
| Final Extension | 72℃ | 10 minutes | 1 cycle |

*Table 3 Cycle parameters for KAPA HiFi HotStart*

2.4.4 PCR Product Digestion + Ligation

Both the PCR product the full length SYCP2 gene and the target plasmid pEGFP-N1 was digested with Promega BamHI (Cat. #R6021) + SalI (Cat. # R6051). Following manufacturer's suggested conditions. Digested Plasmid + Gene of interested were ligated in

a 3:1 plasmid to gene of interest concentration using Invitrogen Anza™ T4 DNA Ligase Master Mix (Cat. # IVGN2104).

### 2.4.5 Plasmid Construct and Bacterial Transformation

Successfully ligated product was transformed into in-house chemically treated (Calcium Chloride) competent cells; 50 uL of thawed competent cells were mixed with 5uL of ligated plasmids gently mixed by flicking tubes, incubated on ice for 30 minutes and then heat shocked in 42℃ for 45 seconds and then added to prewarmed 37℃ SOC media left in shaking incubator for 1 hour and after that plated on Kanamycin antibiotic supplied Agar plates. Plasmids purified as steps shown in 3.3.1.

## 2.5 Immunohistochemistry (IHC)

Fixing of mouse testis tubules was done following (Namekawa and Lee, 2011) Mouse was killed and testis placed in 1x PBS at room temperature, an incision made in testis, and the tubules squeezed into the fixative without letting go of the tunica aluginea. Some tubules (5-10) transferred to CSK buffer with Triton x100 and incubated for 6 mins at room temperature. This step is important for removal of background. Tubules transferred into 4% PFA and incubated for 10 mins and then transferred into fresh PBS and incubated for 5 mins. Tubules placed onto 20-30 uL of PBS on the back of a glass slide then tubules are

teared between two watchmaker forceps for 10-20 secs until homogenised. The tubules are transferred to a Eppendorf tube with 1.3 mL PBS.100 uL of suspension was loaded onto a cytospin chamber and spun at 200 rpm for 10 mins at room temperature. After that slides are washed with PBS for 5 min in a coplin jar and transferred and stored in 70% EtOH.

Immunostaining of mouse testis tubules was done following ( Xu H, Tong Z, Ye Q, et al. 2019)(Xu *et al.*, 2019)

Slides washed in 1x PBS and then blocked with PBT [1% BSA + 0.1% Tween-20 in 50 mL PBS] for 20 min at room temperature. Slides incubated in primary antibodies concentration 1:400 (Millipore Anti-SYCP2 - Polyclonal Antibody. Cat. #ABE2622) in PBT overnight at 4℃. Slides washed in 0.1% Tween-20 in PBS for 5 mins at room temperature followed by incubation of secondary antibodies in PBT for 2 hr at room temperature. This was done in the dark because secondary Antibodies were labelled with fluorophores. And Finally mounted with DAPI and visualised in Fluorescent Microscope.

# 3 Results

## 3.1 Bioinformatic Analysis

Normal expression of SYCP2 is almost completely restricted to testes. To assess the expression profile of SYCP2 across normal tissues and cancer tissue, two different bioinformatic analyses were performed. The first was using GTEx (Genotype Tissue Expression) portal which is a public access database providing tissue specific gene expression analysis. Using its visual interactive genomics tools we were able to obtain a graph showing SYCP2 normal gene expression across different human tissues. (Figure 8) The graph shows a comparison of expression throughout the different human tissues, we can see there is a clear preference of expression to the testis.

The second bioinformatic analysis performed was using CBioPortal, (J *et al.*, 2013) (E *et al.*, 2012) which is a public access interactive bioinformatic tool analysing transcriptional cancer genomics data sets. There is an overwhelming amount of genomics data available online and a tool like CBioPortal allows for easy access and interpretation of molecular profiles of different cancer genomic data sets. The first graph (Figure 9) was done using data from the global initiative; The Cancer Genome Atlas (TCGA) (Weinstein *et al.*, 2013) a large-scale collaboration that provides a comprehensive database of the most prevalent cancers using tumour tissue samples from different cancer patients. The provided data by this initiative allows for study comparisons and evaluation of different cancer types,

exploring the genomic aberrations allows for a better understanding of cancer progression and subsequently possible clinical implication.

All the Pan Cancer Atlas samples in TCGA were evaluated, 32 studies, 10967 samples. The plot (Figure 9) shows the level of mRNA expression (vertical axis) and cancer type (horizontal axis). The plot was sorted by median; we can observe that the highest expression of SYCP2 is in cervical cancer followed by breast invasive carcinoma and so on. But we can see there is not a significant difference in the level of expression between the different cancer tissues.

These findings show that the gene is acting in a noncanonical manner leading to an investigation to look into the mutations in the queried gene (Figure 10). As can be shown the alteration in the gene is mostly an upregulation.


*Figure 8 SYCP2 expression across normal tissue.*

*Data is obtained using GTEx* (*GTEx Portal*, no date) *which is a public data resource set up by the NIH to study gene expression. The graph shows the expression level of SYCP2 through different normal human tissue. Although there is some level of expression in other tissues, the testes show the highest expression.*

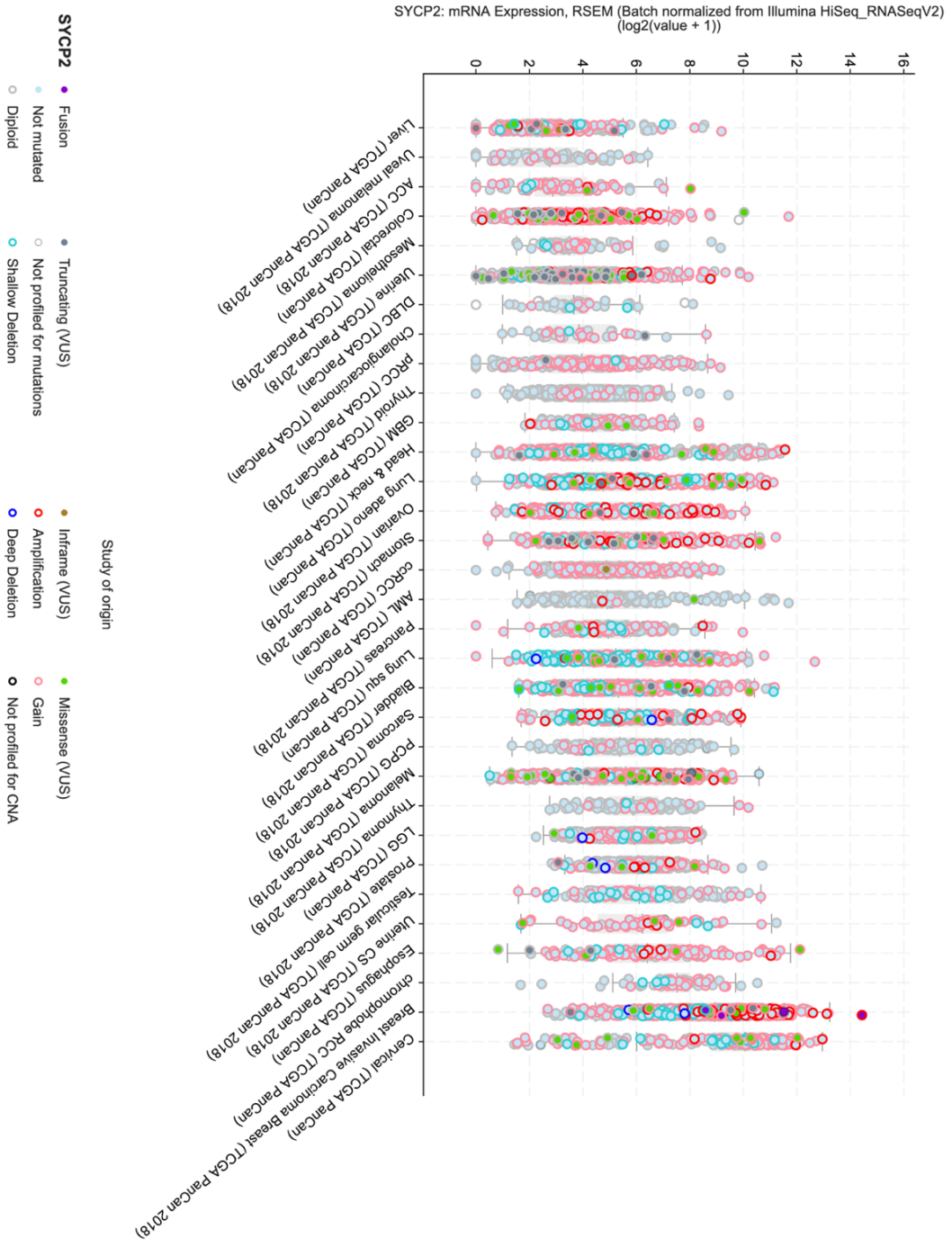


*Figure 9 CBioPortal*(CBioPortal, 2013) *plot showing SYCP2 expression across different cancers.]Data obtained using CBioPortal which is a public access tool providing large scale cancer genomic data sets. This graph shows the expression profile of SYCP2 across different cancers. Plot sorted by median; Showing the highest expression level of SYCP2 is in cervical cancer. Comparing this graph to figure 8 there is a notable significant upregulation of SYCP2 in cancer tissues when compared to their normal tissue. Although the plot very clearly shows significant expression of SYCP2 there is not a --- pattern to their aberrant expression as shown in the differently coloured circles, some are fusion mutations, some are deletions, some are amplification etc.*

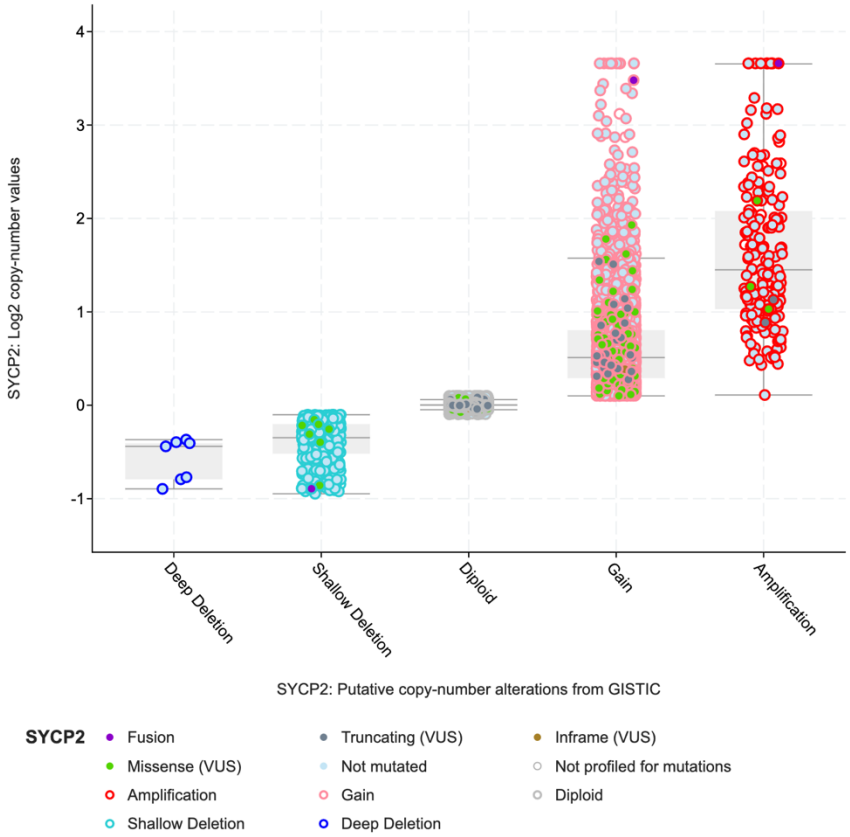


*Figure 10 CBioPortal plot types of alterations of target gene SYCP2.*

*As shown previously, there are different types of mutations associated with SYCP2's wrongful expression in cancer tissues. This plot illustrates that gain of function and amplification mutations are the most found aberrations in SYCP2 when it comes to its non-canonical expression in cancer tissue.*

## 3.2 Western Blots

### 3.2.1 Optimising Western blotting

Existing data shows SYCP2 being expressed on the mRNA, in order to investigate whether the expression on the mRNA is followed by expression on the protein level; Western blotting was performed. Testing both HPV-positive and HPV-negative cervical cancer cell lines with mouse testis lysate as a positive control. The predicted size of the gene of interest SYCP2 is ~200 kDa.

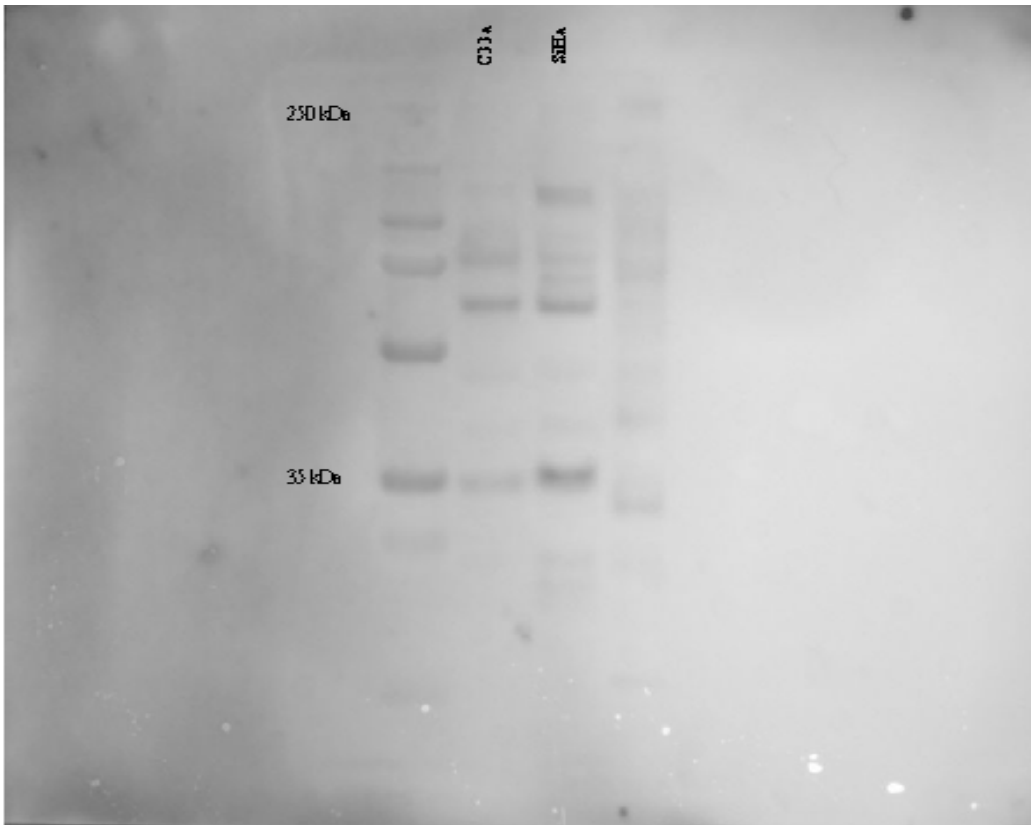


*Figure 11 Western blot showing different protein lysates of HPV+ (SiHa) and HPV- (C33a) and mouse testis lysate as a positive control. The predicted size of SYCP2 is ~200 kDa. SYCP2 Ab. Conc. 1:1000 overnight incubation. Block 5% skimmed milk in TBST 0.1% Tween.*

The first blot performed was obtained using the BioRad Trans-Blot Turbo System as shown in (Figure 11). The semi-dry transfer yielded faint uncharacterised bands. Across all organisms the cell has a host of proteolytic enzymes which could interfere with the integrity of proteins and cause degradation.(Farady and Craik, 2010) To address that and test whether the bands witnessed are degradation products or in fact uncharacterised bands the cells were treated with a proteosome inhibitor.

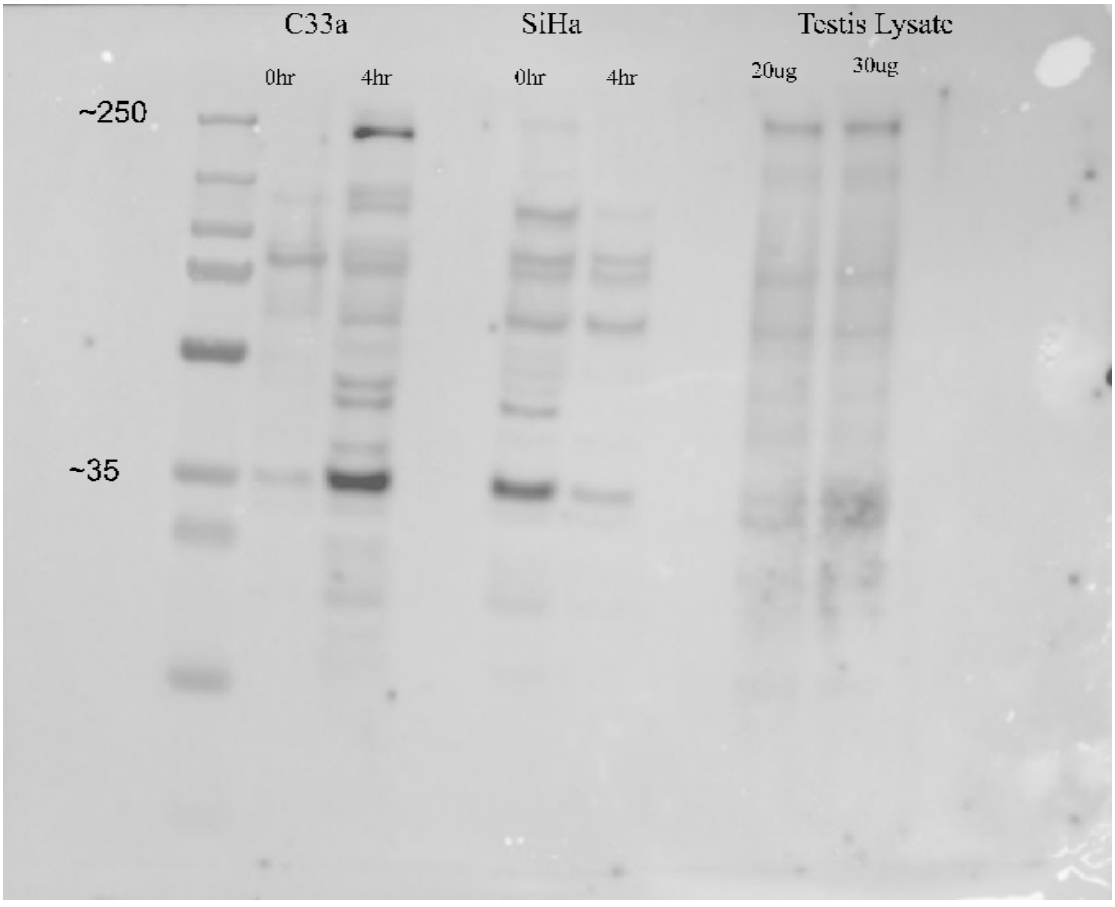


*Figure 12 Western blot showing different protein lysates of HPV+ (SiHa) and HPV- (C33a) treated with MG132 in a time dependent manner [0hr, 30 min, 2hr, 3hr, 4hr] and mouse testis lysate as a positive control. The predicted size of SYCP2 is ~200 kDa. SYCP2 Ab. Conc. 1:1000 overnight incubation. Block 5% skimmed milk in TBST 0.1% Tween.*

The second blot shows the samples after treating with M132 the proteasome inhibitor (Figure 12) but as can be seen, there are still uncharacterised bands. And so to optimise transfer it has been shown in studies (Kong, Pu and Ma, 2008) that wet tank method of transfer yields better results of larger molecular weight proteins and so a wet transfer was attempted using Towbin buffer (Towbin, Staehelin and Gordon, 1979). As shown in (Figure 12) the transfer yielded clearer bands but still not at the predicted size. In further attempts to optimise western blotting and eliminate uncharacterized bands (Figure 13) and (Figure 14) blots were incubated at 2 hours instead of overnight and washing steps extended. But as is clearly visible one uncharacterised band remains in all blots at ~35 kDa.

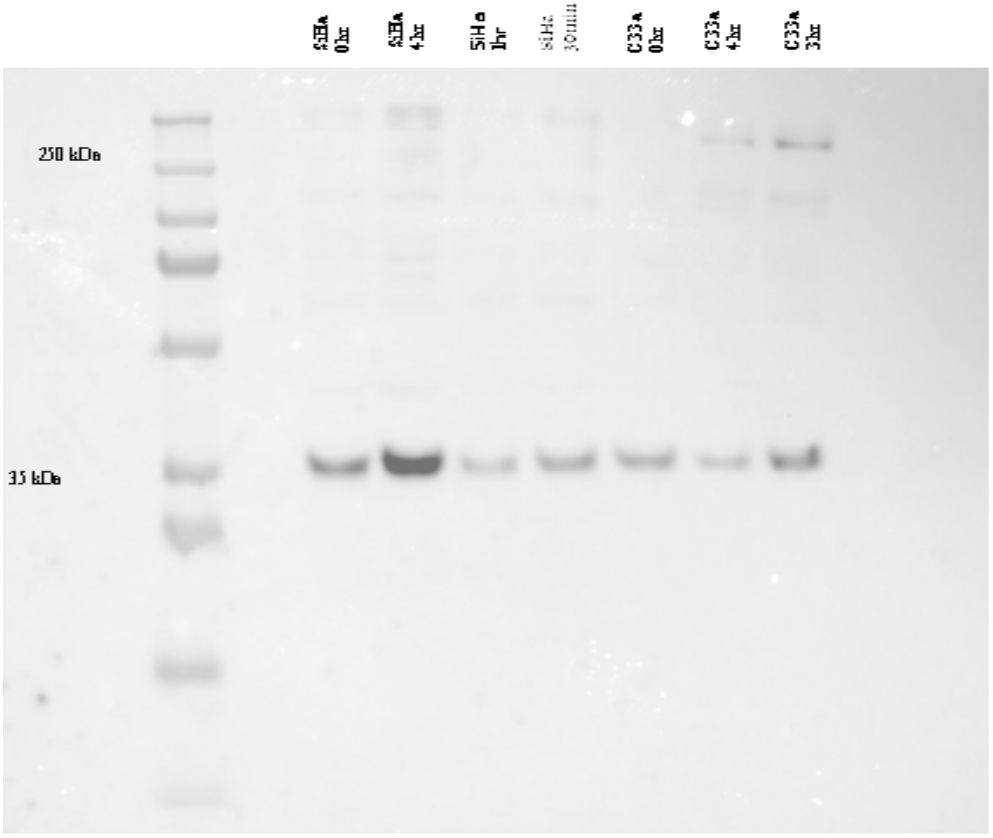


*Figure 13 Western blot showing different protein lysates of HPV+ (SiHa) and HPV- (C33a) treated with MG132 in a time dependent manner [0hr, 30 min, 2hr, 3hr, 4hr] and mouse testis lysate as a positive control. The predicted size of SYCP2 is 200 kDa. SYCP2 Ab. Conc. 1:1000 2hr incubation. Block 5% skimmed milk in TBST 0.1% Tween. Washing steps are also done in TBST 0.1% Tween. Exposure 49 seconds.*

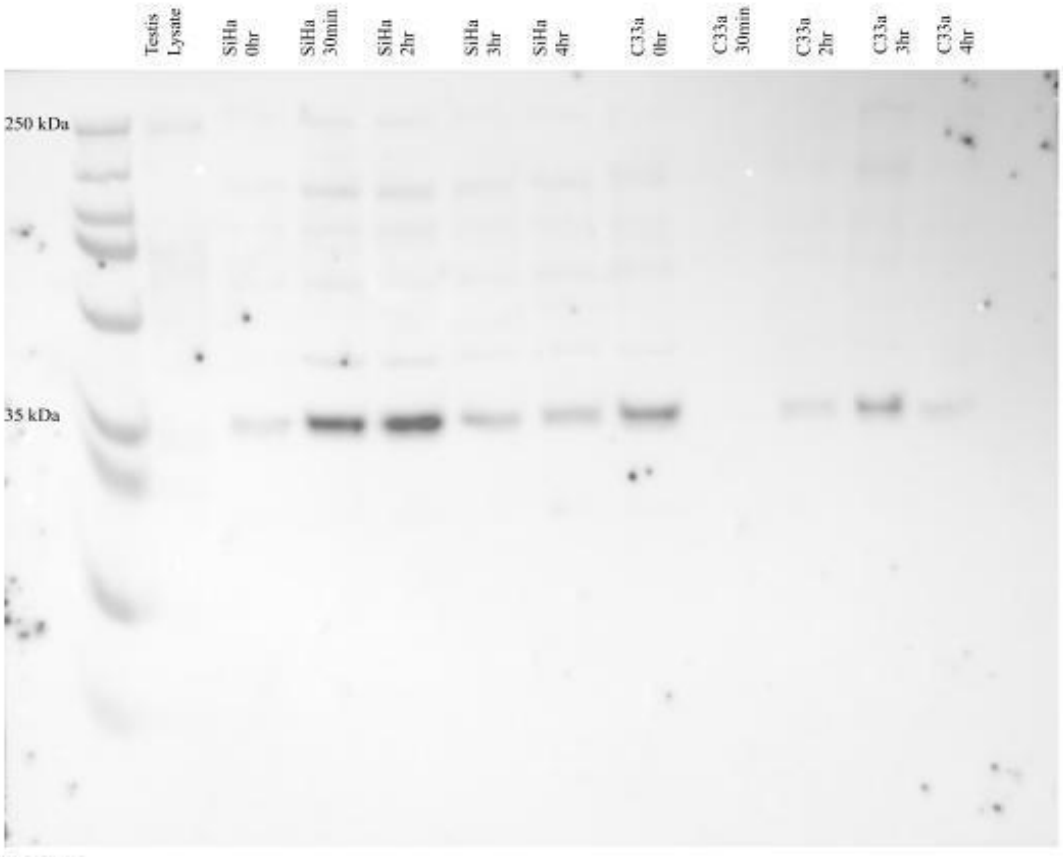


*Figure 14 Western blot showing different protein lysates of HPV+ (SiHa) and HPV- (C33a) treated with MG132 in a time dependent manner [0hr, 30 min, 2hr, 3hr, 4hr] and mouse testis lysate as a positive control. The predicted size of SYCP2 is 200 kDa. (SYCP2 Ab. Conc. 1:1000) 2hr incubation. Block 5% skimmed milk in TBST 0.025% triton. Washing steps done also in TBST 0.025% triton. Exposure 3 minutes.*

Lysis buffers used in western blotting usually do a good job in extracting proteins especially cytoplasmic or membrane bound however in this project case as the gene of interest in nucleus based there is a chance that the protein is not being completely solubilised extracted and being left in the pellet. To test that it is not a solubility issue and the protein is not being lost during the protein extraction method and being left in the pellet the insoluble have been tested using Western blotting (Figure 15) and as shown no clear band is visible which eliminates that it is a solubility issue.

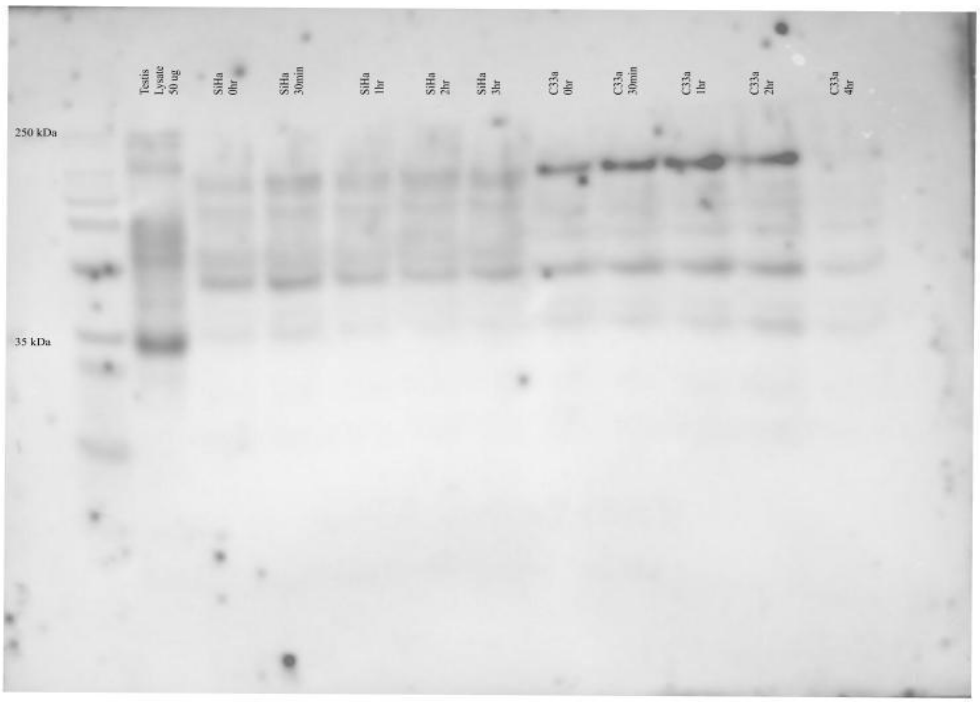


*Figure 15 Western blot showing the MPER-insoluble pellets of HPV+ (SiHa) and HPV- (C33a) treated with MG132 in a time dependent manner [0hr, 30 min, 2hr, 3hr, 4hr] and mouse testis lysate as a positive control. The predicted size of SYCP2 is 200 kDa. Antibody concentration was SYCP2 1:1000 and overnight incubation. The blot shows uncharacterised bands and none of the predicted molecular weight, confirming that the protein is not left behind with insoluble pellets during protein extraction.*

## 3.3 Immunofluorescence Microscopy

To visualise the localised expression of SYCP2 in normal tissue (tissue used here is mouse testis) the same antibody (Millipore Anti-SYCP2 - Polyclonal Antibody. Cat. #ABE2622) The western blotting results observed had some uncharacterised bands, although there were bands in the predicted molecular weight the presence of other bands led to a question of whether or not the purchased antibody is recognising the correct epitope of the gene of interest, because that would hinder the validity of the result. This experiment (figure 16) was conducted to assess if the antibody is recognizing the correct epitope and it confirmed that indeed it is recognising the correct epitope.

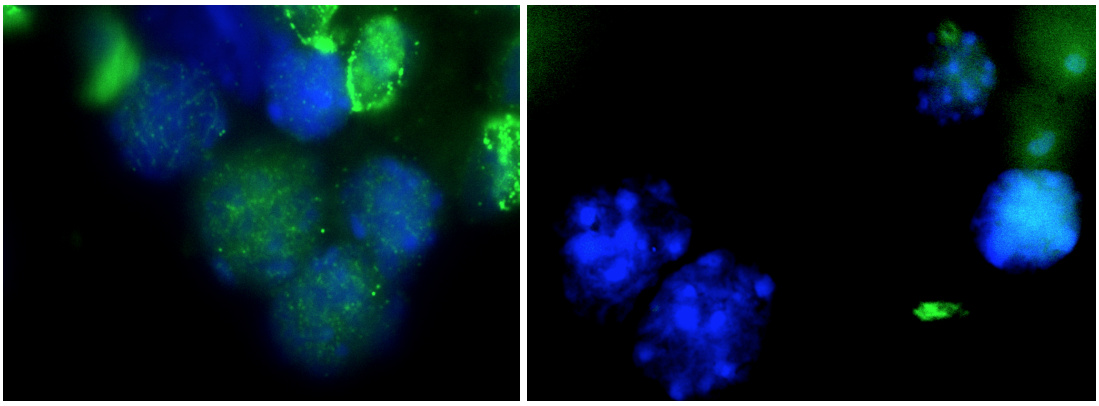

*Figure 16 Mouse testis microtubules with anti-SYCP2 antibody. Left: Positive control SYCP2 concentration of 1:400. Right: Negative control with only secondary antibody. This experiment was conducted to confirm that the purchased antibody was recognising the correct epitope of the gene of interest. To confirm the western blot results observed in (figure14) and (figure15), and eliminate the possibility that the bands observed might be uncharacterised bands.*

## 3.4 Cloning Gene of interest into Plasmid Construct

To validate the western blot results this experiment was performed as the first step to validate protein expression of SYCP2 through assessing a transiently expressed form of the protein from extrachromosomal construct (pEGFP-N1). Visualising the transiently expressed SYCP2 protein on western blot would validate the primary results and would eliminate any questions of solubility, considering that SYCP2 is localised in the nucleus there is still a possibility that it is not being solubilised properly and what was observed is uncharacterised bands. Assessing it using western blotting from extrachromosomal expressed protein would eliminate that possibility.

To express the gene of interest transiently; SYCP2 was amplified from purchased TOPO plasmid expressing the full length gene using HiFi PCR ( Figure 17). Afterwards, the PCR product was digested with BamHI + SalI along with the plasmid construct pEGFP-N1 (Figure18). And ligated using Invitrogen Anza™ T4 DNA Ligase Master Mix. (Figure 19) shows the successfully transformed and isolated construct allowing expression of SYCP2 as a C-terminal GFP-fusion protein.

*Figure 17 KAPA HIFI PCR product. Starting product was 10ng of TOPO plasmid*

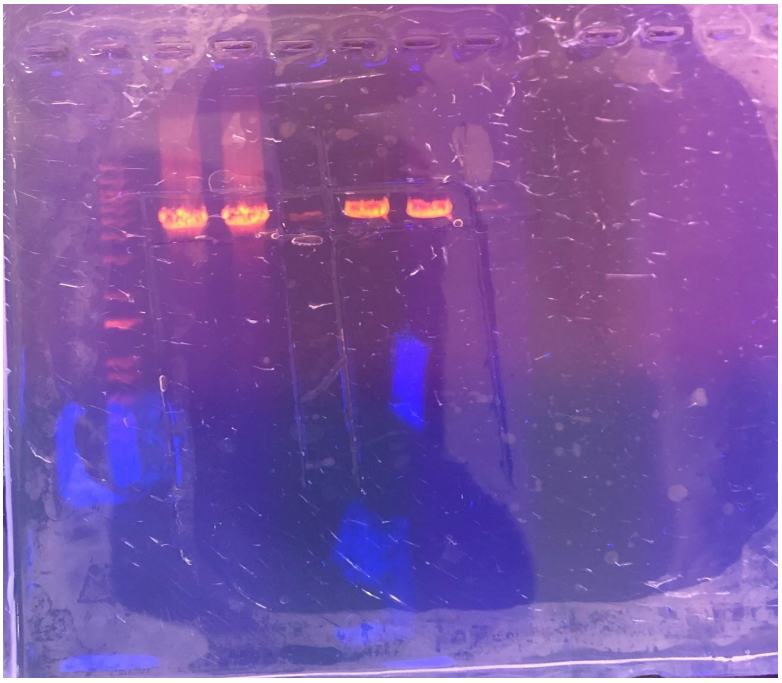

*Figure 18 pEGFP-N1 (third and fourth wells) after restriction digestion with Promega; BamHI and SalI and SYCP2 PCR product (last and second last wells) after restriction digestion with Promega; BamHI and SalI.*

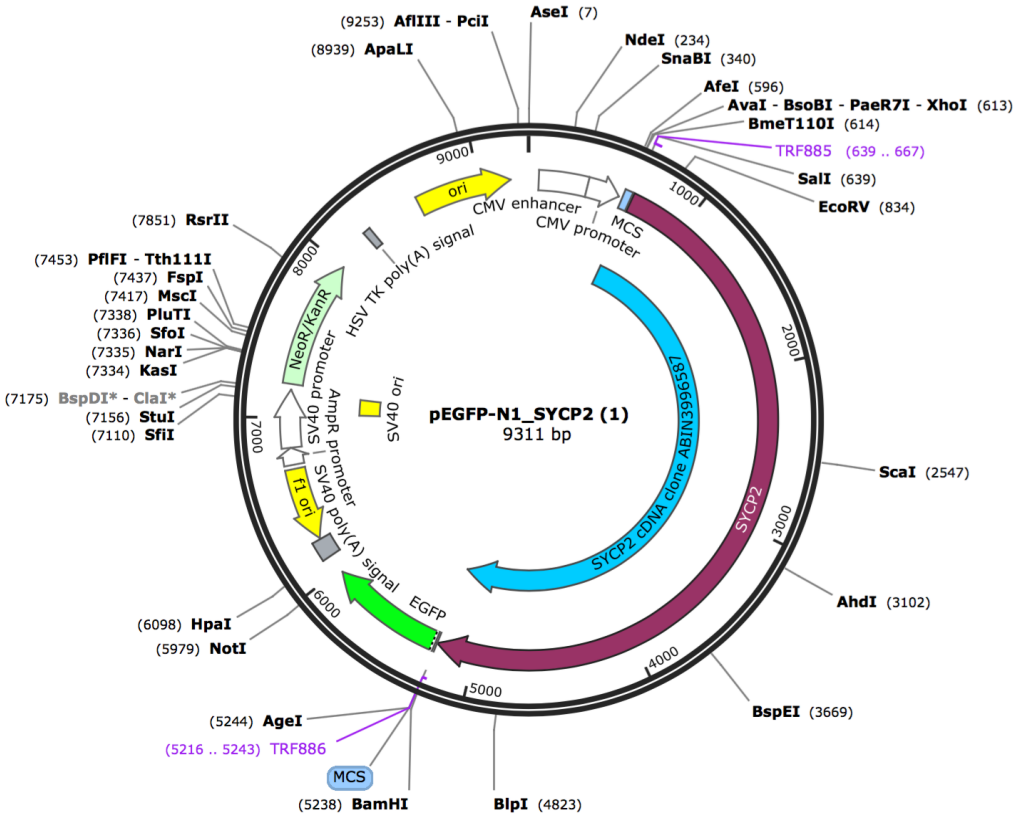


*Figure 19 pEGFP-N1 plasmid expressing full length SYCP2*

# 4 Discussion:

As it has been established in previous studies, SYCP2 is found wrongfully overexpressed in HPV+ cancer cell lines, SYCP2 being part of the synaptonemal complex and the role it has in meiotic recombination if it is expressed outside the controlled environment of meiotically dividing cells, could lead to an environment of genetic instability. Studies conducted showed expression on the mRNA level, to investigate any clinical potential with this finding further experimentation needs to be conducted. Which is what this project set out to investigate; if the expression on the mRNA level is followed by expression on the protein level. Experiments were conducted to assess the aberrant expression of SYCP2 on the protein level in cervical cancer cell lines both HPV+ and HPV-.

There is still very little known in the literature about the aberrant expression of this gene; when normally expressed it aids with the conserved well known process of homologous recombination that happens in meiosis. And it has been explored that not only does it offer a structural role it also plays a role in the choice of recombination pathway in order to accomplish the genetic diversity of progeny which is where DNA amplification differs in meiosis from mitosis in the resulting daughter cells. And by the use of bioinformatic tools we were able to visually compare the difference of expression in normal tissue against the wrongful expression in cancer tumour tissue (Figure 8) (Figure 9). And as can be seen when expressed normally it is almost completely restricted to testes but when expressed aberrantly it can be seen in an array of cancers.

## 4.1 Assessment of SYCP2 Protein Expression

Western blotting was attempted and although it was optimised and better transfer was achieved and more defined bands were observed, there were still uncharacterised bands of an unexpected molecular weight. This could be due to a few reasons; it could be that the protein observed is a degradation product, that was addressed by treating the cells with proteasome inhibitor MG132; which yielded clearer bands but still not at the expected level. Moreover, it has been observed in adenoviruses that while the virus hijacks the host cell machinery, translation of host mRNA is shut down to facilitate the translation of viral proteins. (Schreiner, Wimmer and Dobner, 2012) To address this possibility a transient infection of the cancer cell lines were attempted by purchasing full length SYCP2 gene and successfully inserted into GFP tagged plasmid (Figure 19) to follow that with performing western blotting to assess if the bands observed were due to a molecular event or a defect in the antibody. This was limited due to COVID19 pandemic and subsequent lockdowns. The GFP tagged plasmid would have been first digested with BamHI and SalI to confirm the plasmid is with the correct gene insert and then followed with transient transfection. And the process of western blotting would be achieved subsequent to that.

Because of the localization of the protein in the nucleus it could be that the lysis buffer cannot completely solubilise the proteins and the protein of interest is left behind in the pellet. Although this has been addressed by performing western blotting on the insoluble (pellets) and yielded no bands in the expected molecular weight, it still cannot be

completely excluded as the pellet is a mixture of cellular debris and could agglutinate to form complex structure post lysis not allowing for proper separation and thus proper recognition of the antibody. This could be addressed in future work by using nuclear protein extraction which digests cytosolic proteins following with hypotonic buffer allowing the nucleus to swell and homogenised to allow nuclear proteins to be extracted.

The antibody used in the experiments is a polyclonal antibody, the uncharacterised bands observed could be due to the fact that a polyclonal antibody although would bind to the correct epitope of the antigen of gene of interest but could also recognise and bind to other epitopes (Lipman *et al.*, 2005) as opposed to monoclonal antibodies famously discovered and awarded Nobel prize Köhler and Milstein 1975 (Köhler and Milstein, 1975) where splenic B cells and myeloma cells are fused producing hybridomas producing unique specific monoclonal antibodies with higher affinity would specifically bind to the gene of interest. In an attempt to address that; Immunofluorescence microscopy was conducted with the same antibody to explore whether or not it is in fact attaching to the right epitope. And indeed, the results showed that it is picking up the SYCP2 signal but there is a significant amount of background and especially with the acrosome (Figure 20) that it is clearly picking up a signal for.

More interestingly, demonstrated in (Ahmad *et al.*, 2005) witnessing a difference in observed molecular weight to the predicted molecular weight could mean a post translational modification event is occurring such as alternative splicing.

4.1.1 Human papillomavirus regulation of gene expression effect on host cell genes.

HPV depends extensively on the process of alternative splicing (Johansson and Schwartz, 2013) (Graham and Faizo, 2017) given its small genome to produce its repertoire of viral proteins. Recent advances in RNA sequencing allowed research to be conducted to look into what viral infection effect can have on host splicing (Ashraf *et al.*, 2019) in the study there is a significant amount of host genes showing altered mRNA splicing post viral infection. This molecular event can be due to the virus altering the host splicing machinery as described in 4.1 or a host immune response to the infection (Chang and Zhang, 2017).

## 4.2 SYCP2 possible role oncogenic molecular events

It is unknown why SYCP2 is switched on in HPV cancers. It has been described that HPV favours one DNA repair pathway to another, and as the SC works the same way for meiosis it could be that it is switched on to help the viral infection and production of its progeny. In line with what was found in western blot results, if indeed the SYCP2 protein is alternatively spliced then it could have a completely different function and effect in either viral infection or oncogenesis.

### 4.2.1 Alternative Lengthening of Telomeres

Telomeres are stretches of DNA repetitive (TTAGGG) that cap and protect the chromosomes during replication and maintain genomic integrity such as preventing them from fusing with other chromosomes. With every cell cycle the telomeres shorten, and eventually trigger cellular senescence. This is to maintain the number of cell divisions a cell can undergo. One way for a cell to become a cancer cell is undergoing genetic modifications and entering what is called a “crisis” state having escaped cellular senescence. One mutation is upregulation in the human TERT gene (*hTERT*) which expresses the enzyme telomerase and initiates uncontrolled proliferation of telomeric DNA which is a known hallmark of cancer. However with some tumours cells finds a way to maintain telomeric DNA, independent of telomerase an alternative pathway called; Alternative Lengthening of Telomeres (ALT) the pattern of the telomeric DNA suggests the

ALT cells using the HR pathway using a DNA template activating homologous recombination pathway for DNA repair activated by RAD51 for maintaining the telomeres. The cells are heterogeneous and there are fluctuations in the length of the telomere in each cell and chromosome fusion in some. If SYCP2 is creating neoplastic conditions; considering the role the SC has in meiotic recombination, this molecular process could be how SYCP2 is aiding oncogenesis.

### 4.2.2 What wrongful expression of SYCP2 means to oncogenic progression

As previously discussed in 1.2.1 the activation of meiotic genes in somatic cells are factors leading to progression of cancer. This is due to the unique characteristics meiosis specific genes have that when wrongfully expressed function as oncogenic drivers, such as favouring interhomolog repair which leads to loss of heterozygosity, the genesis of DSB, and alterations of the transcriptional landscape and epigenetic modification (Feichtinger and McFarlane, 2019). This study aimed to assess whether the expression of SYCP2 on the mRNA level is followed by expression on the protein level, when explored via bioinformatic tools we can clearly see the difference in expression when comparing the mRNA expression levels of SYCP2 in normal tissue against tumour tissue as observed in (Figure 8), (Figure 9). Furthermore, barring some difficulties faced with wet lab approach we can confidently attest there is expression being picked up by antibodies against the SYCP2 protein. This proposes an epigenetic alteration module of SYCP2 overexpression in HPV+ cancers.

## 4.3 Conclusions and Future Prospects

HPV makes a unique model because we can readily examine the changes on the molecular level in what it could mean to oncogenesis. It has been shown that SYCP2 is overexpressed and whilst difficulties were faced in the assessment of protein expression, further experimentations can investigate the possibility of alternative splice events using combination of western blotting and transcript identification by rapid amplification of cDNA ends (RACE) to assess which splice forms of SYCP2 are expressed in a panel of HPV16+ cancer cell lines. And to investigate if SYCP2 is partaking a role in HPV infection, using CRISPR-Cas9 to delete SYCP2 in normal keratinocytes (NIKS) will allow investigating if HPV16 is able to complete genome replication and amplification in organotypic epithelia generated from SYCP2-KO cells. A question remains if the results found are experimental shortcoming or a finding that there is a splicing event, although this finding does not demonstrate an active role of this gene in cancer progression the data warrants future investigations.

Further research needs to be conducted to investigate the role SYCP2 could have, if it is switched on after the aborted cycle of HPV and lead to oncogenesis or if it is switched on early on to aid HPV in its viral replication. Or if that it is simply a by-product of the various molecular changes HPV cancers present that have no substantial role in viral infection or oncogenesis.

# 5 References

Ahmad, Q. R. *et al.* (2005) 'Molecular weight assessment of proteins in total proteome profiles using 1D-PAGE and LC/MS/MS', *Proteome Science*. BioMed Central, 3, p. 6. doi: 10.1186/1477-5956-3-6.

Alberts, B. *et al.* (2019) *Essential Cell Biology: Fifth International Student Edition*.

Araldi, R. P. *et al.* (2017) 'Papillomaviruses: A systematic review', *Genetics and Molecular Biology*. Brazilian Journal of Genetics, pp. 1–21. doi: 10.1590/1678-4685-GMB-2016-0128.

Araldi, R. P. *et al.* (2018a) 'The human papillomavirus (HPV)-related cancer biology: An overview', *Biomedicine and Pharmacotherapy*. Elsevier, 106(April), pp. 1537–1556. doi: 10.1016/j.biopha.2018.06.149.

Araldi, R. P. *et al.* (2018b) 'The human papillomavirus (HPV)-related cancer biology: An overview', *Biomedicine and Pharmacotherapy*. Elsevier Masson SAS, pp. 1537–1556. doi: 10.1016/j.biopha.2018.06.149.

Ashraf, U. *et al.* (2019) 'Advances in Analyzing Virus-Induced Alterations of Host Cell Splicing', *Trends in Microbiology*. Elsevier Ltd, pp. 268–281. doi:

10.1016/j.tim.2018.11.004.

Van Baren, N. *et al.* (2005) 'Tumoral and immunologic response after vaccination of melanoma patients with an ALVAC virus encoding MAGE antigens recognized by T cells', *Journal of Clinical Oncology*. J Clin Oncol, 23(35), pp. 9008–9021. doi: 10.1200/JCO.2005.08.375.

Boël, P. *et al.* (1995) 'BAGE: a new gene encoding an antigen recognized on human melanomas by cytolytic T lymphocytes', *Immunity*. Immunity, 2(2), pp. 167–175. doi: 10.1016/S1074-7613(95)80053-0.

Bosch, F. X. *et al.* (2002) 'The causal relation between human papillomavirus and cervical cancer', *Journal of Clinical Pathology*. BMJ Publishing Group, pp. 244–265. doi: 10.1136/jcp.55.4.244.

Van Der Bruggen, P. *et al.* (1991) 'A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma', *Science*. Science, 254(5038), pp. 1643–1647. doi: 10.1126/science.1840703.

Buck, C. B., Day, P. M. and Trus, B. L. (2013) 'The papillomavirus major capsid protein L1', *Virology*. Virology, 445(1–2), pp. 169–174. doi: 10.1016/j.virol.2013.05.038.

Burgos-Panadero, R. *et al.* (2019) 'The tumour microenvironment as an integrated framework to understand cancer biology', *Cancer Letters*. Elsevier

Ireland Ltd, pp. 112–122. doi: 10.1016/j.canlet.2019.07.010.

Cai, Q. *et al.* (2013) 'Human papillomavirus early proteins and apoptosis', *Archives of Gynecology and Obstetrics*. Springer, 287(3), pp. 541–548. doi: 10.1007/s00404-012-2665-z.

Castellsagué, X. (2008) 'Natural history and epidemiology of HPV infection and cervical cancer', *Gynecologic Oncology*. Elsevier, 110(3 SUPPL.2), pp. S4–S7. doi: 10.1016/j.ygyno.2008.07.045.

CBioPortal (2013) 'cBioPortal for Cancer Genomics', pp. 2014–2016. Available at: https://www.cbioportal.org/ (Accessed: 29 October 2020).

Chakravarthy, A. *et al.* (2016) 'Human papillomavirus drives tumor development throughout the head and neck: Improved prognosis is associated with an immune response largely restricted to the Oropharynx', *Journal of Clinical Oncology*, 34(34), pp. 4132–4141. doi: 10.1200/JCO.2016.68.2955.

Chakravarthy, A. *et al.* (2020) 'Pro-metastatic gene expression, immune evasion and an altered HPV spectrum characterize an aggressive subtype of cervical cancer', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.04.02.019711. doi: 10.1101/2020.04.02.019711.

Chang, M. and Zhang, J. (2017) 'Alternative Pre-mRNA Splicing in Mammals and Teleost Fish: A Effective Strategy for the Regulation of Immune

Responses Against Pathogen Infection', *International Journal of Molecular Sciences*. MDPI AG, 18(7), p. 1530. doi: 10.3390/ijms18071530.

Chen, Y. T. *et al.* (1997) 'A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 94(5), pp. 1914–1918. doi: 10.1073/pnas.94.5.1914.

Chow, L. T., Broker, T. R. and Steinberg, B. M. (2010) 'The natural history of human papillomavirus infections of the mucosal epithelia', *APMIS*. APMIS, pp. 422–449. doi: 10.1111/j.1600-0463.2010.02625.x.

Costa, F. F., Le Blanc, K. and Brodin, B. (2007) 'Concise Review: Cancer/Testis Antigens, Stem Cells, and Cancer', *STEM CELLS*. Wiley, 25(3), pp. 707–711. doi: 10.1634/stemcells.2006-0469.

Day, P. M. and Schelhaas, M. (2014) 'Concepts of papillomavirus entry into host cells', *Current Opinion in Virology*. Elsevier B.V., pp. 24–31. doi: 10.1016/j.coviro.2013.11.002.

Van Doorslaer, K. (2013) 'Evolution of the Papillomaviridae', *Virology*, 445(1–2), pp. 11–20. doi: 10.1016/j.virol.2013.05.012.

E, C. *et al.* (2012) 'The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data', *Cancer discovery*. Cancer

Discov, 2(5), pp. 401–404. doi: 10.1158/2159-8290.CD-12-0095.

Van den Eynde, B. *et al.* (1995) 'A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma', *Journal of Experimental Medicine*. J Exp Med, 182(3), pp. 689–698. doi: 10.1084/jem.182.3.689.

Farady, C. J. and Craik, C. S. (2010) 'Mechanisms of Macromolecular Protease Inhibitors', *ChemBioChem*. NIH Public Access, pp. 2341–2346. doi: 10.1002/cbic.201000442.

Faraji, F. *et al.* (no date) 'Molecular mechanisms of human papillomavirus-related carcinogenesis in head and neck cancer'. doi: 10.1016/j.micinf.2017.06.001.

Feichtinger, J. and McFarlane, R. J. (2019) 'Meiotic gene activation in somatic and germ cell tumours', *Andrology*. Wiley-Blackwell, 7(4), p. 415. doi: 10.1111/ANDR.12628.

Fratta, E. *et al.* (2011) 'The biology of cancer testis antigens: Putative function, regulation and therapeutic potential', *Molecular Oncology*, 5(2), pp. 164–182. doi: 10.1016/j.molonc.2011.02.001.

Geisinger, A. and Benavente, R. (2017) 'Mutations in Genes Coding for Synaptonemal Complex Proteins and Their Impact on Human Fertility',

*Cytogenetic and Genome Research*, 150(2), pp. 77–85. doi: 10.1159/000453344.

Gibbs, Z. A. and Whitehurst, A. W. (2018) 'Emerging Contributions of Cancer/Testis Antigens to Neoplastic Behaviors', *Trends in Cancer*. Cell Press, pp. 701–712. doi: 10.1016/j.trecan.2018.08.005.

Gillison, M. L. *et al.* (2019) 'Human papillomavirus and the landscape of secondary genetic alterations in oral cancers'. doi: 10.1101/gr.241141.118.

Graham, S. V. and Faizo, A. A. A. (2017) 'Control of human papillomavirus gene expression by alternative splicing', *Virus Research*. Elsevier B.V., 231, pp. 83–95. doi: 10.1016/j.virusres.2016.11.016.

*GTEx Portal* (no date). Available at: https://www.gtexportal.org/home/gene/ENSG00000196074 (Accessed: 29 October 2020).

Guo, T. *et al.* (2020) 'Chromatin structure regulates cancer-specific alternative splicing events in primary HPV-related oropharyngeal squamous cell carcinoma', *Epigenetics*. Taylor & Francis, 15(9), pp. 959–971. doi: 10.1080/15592294.2020.1741757.

Gutierrez-Xicotencatl, L. *et al.* (2020) 'Cellular functions of HPV16-E5 oncoprotein during oncogenic transformation', *Molecular Cancer Research*.

Mol Cancer Res, p. molcanres.0491.2020. doi: 10.1158/1541-7786.MCR-20-0491.

Helfer, C., Yan, J. and You, J. (2014) 'The Cellular Bromodomain Protein Brd4 has Multiple Functions in E2-Mediated Papillomavirus Transcription Activation', *Viruses*. Multidisciplinary Digital Publishing Institute, 6(8), pp. 3228–3249. doi: 10.3390/v6083228.

*Human papillomavirus (HPV) and cervical cancer* (no date). Available at: https://www.who.int/en/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer (Accessed: 22 October 2020).

J, G. *et al.* (2013) 'Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal', *Science signaling*. Sci Signal, 6(269). doi: 10.1126/SCISIGNAL.2004088.

Jang, M. K., Shen, K. and McBride, A. A. (2014) 'Papillomavirus Genomes Associate with BRD4 to Replicate at Fragile Sites in the Host Genome', *PLoS Pathogens*. Edited by P. F. Lambert. Public Library of Science, 10(5), p. e1004117. doi: 10.1371/journal.ppat.1004117.

Johansson, C. and Schwartz, S. (2013) 'Regulation of human papillomavirus gene expression by splicing and polyadenylation', *Nature Reviews Microbiology*. Nat Rev Microbiol, pp. 239–251. doi: 10.1038/nrmicro2984.

Kawai, A. *et al.* (1998) 'SYT-SSX gene fusion as a determinant of morphology and prognosis in synovial sarcoma', *New England Journal of Medicine*, 338(3), pp. 153–160. doi: 10.1056/NEJM199801153380303.

KH, V. and DP, L. (2007) 'p53 in health and disease', *Nature reviews. Molecular cell biology*. Nat Rev Mol Cell Biol, 8(4), pp. 275–283. doi: 10.1038/NRM2147.

Köhler, G. and Milstein, C. (1975) 'Continuous cultures of fused cells secreting antibody of predefined specificity', *Nature*. Nature Publishing Group, 256(5517), pp. 495–497. doi: 10.1038/256495a0.

Kong, L. Q., Pu, Y. H. and Ma, S. K. (2008) 'Better performance of Western blotting: quick vs slow protein transfer, blotting membranes and the visualization methods', *Nan fang yi ke da xue xue bao = Journal of Southern Medical University*. Nan Fang Yi Ke Da Xue Xue Bao, 28(1), pp. 26–29. Available at: https://pubmed.ncbi.nlm.nih.gov/18227019/ (Accessed: 17 October 2020).

Lechner, M. and Fenton, T. R. (2016) 'The Genomics, Epigenomics, and Transcriptomics of HPV-Associated Oropharyngeal Cancer-Understanding the Basis of a Rapidly Evolving Disease', *Advances in Genetics*. Academic Press Inc., 93, pp. 1–56. doi: 10.1016/bs.adgen.2015.12.001.

Li, X. C., Bolcun-Filas, E. and Schimenti, J. C. (2011) ‘Genetic evidence that synaptonemal complex axial elements govern recombination pathway choice in mice’, *Genetics*, 189(1), pp. 71–82. doi: 10.1534/genetics.111.130674.

Lipman, N. S. *et al.* (2005) ‘Monoclonal Versus Polyclonal Antibodies: Distinguishing Characteristics, Applications, and Information Resources’, *ILAR Journal*. Institute for Laboratory Animal Research, 46(3), pp. 258–268. doi: 10.1093/ilar.46.3.258.

Ma, T. *et al.* (1999) ‘Interaction between cyclin-dependent kinases and human papillomavirus replication-initiation protein E1 is required for efficient viral replication’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 96(2), pp. 382–387. doi: 10.1073/pnas.96.2.382.

Masterson, L. *et al.* (2015) ‘Deregulation of SYCP2 predicts early stage human papillomavirus-positive oropharyngeal carcinoma: A prospective whole transcriptome analysis’, *Cancer Science*, 106(11), pp. 1568–1575. doi: 10.1111/cas.12809.

McBride, A. A. (2008) ‘Chapter 4 Replication and Partitioning of Papillomavirus Genomes’, *Advances in Virus Research*. NIH Public Access, pp. 155–205. doi: 10.1016/S0065-3527(08)00404-1.

McBride, A. A. (2017a) ‘Mechanisms and strategies of papillomavirus replication’, *Biological Chemistry*, 398(8), pp. 919–927. doi: 10.1515/hsz-2017-0113.

McBride, A. A. (2017b) ‘Mechanisms and strategies of papillomavirus replication’, *Biological Chemistry*. Walter de Gruyter GmbH, 398(8), pp. 919–927. doi: 10.1515/hsz-2017-0113.

McBride, A. A. (2017c) ‘Oncogenic human papillomaviruses’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1732). doi: 10.1098/rstb.2016.0273.

McBride, A. A. and Warburton, A. (2017) ‘The role of integration in oncogenic progression of HPV-associated cancers’, *PLoS Pathogens*. Public Library of Science. doi: 10.1371/journal.ppat.1006211.

McFarlane, R. J. and Wakeman, J. A. (2017) ‘Meiosis-like functions in oncogenesis: A new view of cancer’, *Cancer Research*, 77(21), pp. 5712–5716. doi: 10.1158/0008-5472.CAN-17-1535.

Moody, C. A. and Laimins, L. A. (2009) ‘Human papillomaviruses activate the ATM DNA damage pathway for viral genome amplification upon differentiation.’, *PLoS pathogens*. Public Library of Science, 5(10). doi: 10.1371/journal.ppat.1000605.

Moody, C. A. and Laimins, L. A. (2010) ‘Human papillomavirus oncoproteins: Pathways to transformation’, *Nature Reviews Cancer*. Nat Rev Cancer, pp. 550–560. doi: 10.1038/nrc2886.

Moore, L. D., Le, T. and Fan, G. (2013) ‘DNA methylation and its basic function’, *Neuropsychopharmacology*. Nature Publishing Group, pp. 23–38. doi: 10.1038/npp.2012.112.

Namekawa, S. H. and Lee, J. T. (2011) ‘Detection of nascent RNA, single-copy DNA and protein localization by immunoFISH in mouse germ cells and preimplantation embryos’, *Nature Protocols*. Nat Protoc, 6(3), pp. 270–284. doi: 10.1038/nprot.2010.195.

nd Kelsey C. MartinMhatre V. Ho, Ji-Ann Lee, A. (2012) ‘The Logic and Mechanism of Homologous Recombination Partner Choice’, *Bone*, 23(1), pp. 1–7. doi: 10.1038/jid.2014.371.

Pardo, B., Gómez-González, B. and Aguilera, A. (2009) ‘DNA double-strand break repair: How to fix a broken relationship’, *Cellular and Molecular Life Sciences*, 66(6), pp. 1039–1056. doi: 10.1007/s00018-009-8740-3.

Reece, J. B. *et al.* (no date) ‘Regulation of Gene Expression’, in *Campbell Biology (9th Edition)*.

Rivlin, N. *et al.* (2011) ‘Mutations in the p53 Tumor Suppressor Gene:

Important Milestones at the Various Steps of Tumorigenesis', *Genes & Cancer*. Impact Journals, LLC, 2(4), p. 466. doi: 10.1177/1947601911408889.

Sahin, U. *et al.* (1995) 'Human neoplasms elicit multiple specific immune responses in the autologous host', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 92(25), pp. 11810–11813. doi: 10.1073/pnas.92.25.11810.

Schreiner, S., Wimmer, P. and Dobner, T. (2012) 'Adenovirus degradation of cellular proteins', *Future Microbiology*. Future Microbiol, pp. 211–225. doi: 10.2217/fmb.11.153.

Serrano, B. *et al.* (2015) 'Human papillomavirus genotype attribution for HPVs 6, 11, 16, 18, 31, 33, 45, 52 and 58 in female anogenital lesions', *European Journal of Cancer*. Elsevier Ltd, 51(13), pp. 1732–1741. doi: 10.1016/j.ejca.2015.06.001.

Shakib, K. *et al.* (2005) 'Proteomics profiling of nuclear proteins for kidney fibroblasts suggests hypoxia, meiosis, and cancer may meet in the nucleus', *Proteomics*, 5(11), pp. 2819–2838. doi: 10.1002/pmic.200401108.

Slebos, R. J. C. *et al.* (2006) 'Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma', *Clinical Cancer Research*, 12(3 I), pp. 701–709. doi:

10.1158/1078-0432.CCR-05-2017.

Stanley, M. A. (2012) 'Epithelial cell responses to infection with human papillomavirus', *Clinical Microbiology Reviews*. American Society for Microbiology (ASM), pp. 215–222. doi: 10.1128/CMR.05028-11.

Towbin, H., Staehelin, T. and Gordon, J. (1979) 'Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 76(9), pp. 4350–4354. doi: 10.1073/pnas.76.9.4350.

Del Val, E. *et al.* (2019) 'RecA and DNA recombination: A review of molecular mechanisms', *Biochemical Society Transactions*, 47(5), pp. 1511–1531. doi: 10.1042/BST20190558.

Wang, H. K. *et al.* (2009) 'Robust production and passaging of infectious HPV in squamous epithelium of primary human keratinocytes', *Genes and Development*. Cold Spring Harbor Laboratory Press, 23(2), pp. 181–194. doi: 10.1101/gad.1735109.

Weinstein, J. N. *et al.* (2013) 'The cancer genome atlas pan-cancer analysis project', *Nature Genetics*. Nature Publishing Group, pp. 1113–1120. doi: 10.1038/ng.2764.

West, A. M. *et al.* (2019) ‘A conserved filamentous assembly underlies the structure of the meiotic chromosome axis’, *eLife*, 8, pp. 1–27. doi: 10.7554/eLife.40372.

Whitehurst, A. W. (2014) ‘Cause and consequence of cancer/testis antigen activation in cancer’, *Annual Review of Pharmacology and Toxicology*. Annual Reviews Inc., pp. 251–272. doi: 10.1146/annurev-pharmtox-011112-140326.

von Witzleben, A. *et al.* (2020) ‘HNSCC: Tumour Antigens and Their Targeting by Immunotherapy’, *Cells*. NLM (Medline). doi: 10.3390/cells9092103.

Xu, H. *et al.* (2019) ‘Molecular organization of mammalian meiotic chromosome axis revealed by expansion STORM microscopy’, *Proceedings of the National Academy of Sciences of the United States of America*, 116(37), pp. 18423–18428. doi: 10.1073/pnas.1902440116.

Yang, F. *et al.* (2006) ‘Mouse SYCP2 is required for synaptonemal complex assembly and chromosomal synapsis during male meiosis’, *Journal of Cell Biology*, 173(4), pp. 497–507. doi: 10.1083/jcb.200603063.