Cindy Dong

Investigating the expression and function of SYCP2 in HPV+ cancer cell lines



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

MSc by Research (Cell Biology)

2019

Supervisors:

Dr Tim Fenton, Dr Peter Ellis

The University of Kent

Faculty of Science

School of Biosciences

Cindy Dong

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.

Cindy Dong

October 2019

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisors, Dr. Tim Fenton and Dr. Peter Ellis. Without their great knowledge, kind support and encouragement, this project would not have been possible. Thanks to my master's degree, I have improved significantly in my laboratory skills and I look forward to a career in science. Also, I would like to thank for all the help and advice that I have got from Max, Nikki and Nerissa who are in Dr Tim Fenton's research group. They have been very patient with my pestering questions throughout this year.

I would also like to thank the University of Kent Korfball club, of which I have been part of since my undergraduate degree. Participating in a sports club this year has helped a lot with stress relief and I am incredibly grateful for all the good friends that I have met.

Lastly, a special thanks to my wonderful family and friends for their constant support and love.

Contents

Declaration **2**
Acknowledgements **3**
Contents **4**
List of Tables **7**
List of Figures **7**
Abbreviations **9**
Abstract **11**
1 Introduction **12**
1.1 Human papillomavirus 12
1.1.1 Structure and genome 12
1.1.2 The infectious cycle of HPV 17
1.1.3 HPV and cancer 21
1.1.3.1 HPV prevalence 21
1.1.3.2 HPV-related cancers 22
1.1.3.3 HPV genome integration 23
1.1.3.4 HPV oncoproteins 24
1.2 Cancer testis antigens 25
1.2.1 CTA expression 25
1.2.2 Functional role of CTAs in cancer 26
1.3 Synaptonemal complex protein 2 28
1.3.1 Meiotic function 28
1.3.2 Overexpression of SYCP2 in HPV-positive cancers 32
1.4 Gene silencing/editing mechanism 34
1.4.1 RNA interference 34
1.4.2 CRISPR/Cas9 genome editing technology 36
1.5 Project outline 39
2 Materials and methods **40**
2.1 Tissue Culture 40
2.1.1 Mammalian cancer cells 40
2.1.2 NIKS 41
2.2 RNA and protein extraction 41
2.3 RNA analysis 43

2.3.1 cDNA synthesis 43
2.3.2 Quantitative PCR 43
2.4 Protein quantification and analysis 45
2.4.1 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) 45
2.4.2 Western blot analysis 45
2.5 Generating standard curve for *SYCP2* quantification 46
2.5.1 Topo cloning 46
2.5.2 Bacterial transformation 47
2.5.3 Liquid culture and plasmid DNA purification 48
2.5.4 SYCP2 standard curve 48
2.6 RNA interference with siRNA 49
2.6.1 Reverse transfection of siRNA 49
2.6.2 Forward transfection of siRNA 50
2.7 CRISPR knockout in NIKS 51
2.7.1 gRNA and vector construct 51
2.7.2 NIKS transfection 53
2.7.3 Flow cytometry 53
2.7.4 DNA extraction 54
2.7.5 Identification of indels with next generation sequencing (NGS) 54
2.7.5.1 Primers design 55
2.7.5.2 PCR amplification 56
2.7.5.3 Generation and purification of amplicons 57
3. Results 58
3.1 Establishment and optimisation of quantification assay for *SYCP2* 58
3.1.1 Successful gene cloning 58
3.1.2 SYCP2 standard curve 60
3.2 Validating SYCP2 overexpression 63
3.3 Effect of siRNA gene silencing on cell viability 65
3.3.1 MTS assay for siRNA transfection efficiency 65
3.3.2 Confirmation of gene silencing via qPCR 69
3.4 Generation of Indels using CRISPR-Cas9 system 71
3.4.1 Purity and sequences of gRNAs 71
3.4.2 FACS 73
3.4.3 Amplification and purification of DNA from edited cell population 75
3.4.4 Next generation sequencing analysis 77
3.4.5 CRISPR analysis 80

4 Discussion.....81
4.1 SYCP2 mRNA is upregulated in HPV16+ cell lines.....81
4.2 Cell viability is not affected by *SYCP2* gene knockdown.....83
4.3 Possible mechanism for SYCP2 expression in HPV+ cancers.....84
4.3.1 Epigenetic regulation.....84
4.3.2 Genomic instability.....86
4.4 Application of CRISPR-Cas9 system in NIKS cells.....89
4.5 Established quantification assay for *SYCP2* mRNA.....92
4.6 Conclusion and future directions.....93
4.6.1 Conclusion.....93
4.6.2 Future directions.....94
5. References.....95

List of Tables

Table 2.1 Mammalian HPV+/HPV- cancer cell lines 40
Table 2.2 Cycle parameters for comparative and standard curve qPCR 44
Table 2.3 Primer sequences for qPCR 44
Table 2.4 Primers designed to use in Platinum SuperFi PCR 46
Table 2.5 siRNAs used in reverse siRNA transfection 50
Table 2.6 *SYCP2* gRNA sequences 51
Table 2.7 Primer designs with universal adaptor and corresponding amplicon sizes 55
Table 2.8 KAPA PCR mixture 56
Table 2.9 PCR cycle parameters 56
Table 3.1 Linear equation of standard curve for targeted genes and primer efficiency 62
Table 3.2 DNA concentration and purity of Horizon SYCP2 gRNAs miniprep 71

List of Figures

Figure 1.1 The phylogeny of HPV 13
Figure 1.2 HPV16 genome structure 14
Figure 1.3 Structure of synaptonemal complex 29
Figure 1.4 Mechanism of RNAi 35
Figure 1.5 Mechanism of CRISPR-Cas9 gene editing technology 38
Figure 2.1 Map of pCR-Blunt II-Topo 47
Figure 2.2 Linear map of *SYCP2* genome with gRNA positions 51
Figure 2.3 Genome map of pD1301-AD with *SYCP2* gRNA insert 52
Figure 2.4 Sequence of overhang universal adaptors 55
Figure 3.1 Gel electrophoresis of miniprep samples resulted from Topo PCR cloning .. 58
Figure 3.2 Illustration of partial *SYCP2* sequence and corresponding sequencing data 59
Figure 3.3 Standard curve for SYCP2 qPCR primers 60
Figure 3.4 Amplification and melt curve plot for miniprep sample using qPCR 61
Figure 3.5 Expression of SYCP2 and E6 relative to TBP in RNA samples extracted from HPV+/HPV- cancer cell lines 64
Figure 3.6 Microscopic images of siRNA transfected UM-SCC-104 cells on day 2 and day 5 post transfection 66
Figure 3.7 Microscopic images of PLK1 transfected 93-VU-147T and SiHa cells on day 5 post transfection 67
Figure 3.8 Percentage of viable cells post-siRNA transfection using MTS assay 67

Figure 3.9 Relative SYCP2 expression in three siRNA transfected HPV+ cell lines 69

Figure 3.10 Relative SYCP2 expression level in HPV+/HPV- cell lines and siRNA transfected HPV16+ cell lines 70

Figure 3.11 Chromatogram data of gRNA sequences 72

Figure 3.12 Evaluation of CRISPR transfection in NIKS under fluorescence microscope and assigned gates in FACS 73

Figure 3.13 Gel electrophoresis of amplicon products from DNA amplification 75

Figure 3.14 Gel electrophoresis of amplicon products from wt NIKS 76

Figure 3.15 Gel electrophoresis of purified amplicons 76

Figure 3.16 Gene editing efficiency of gRNAs 77

Figure 3.17 Illustration of the most common indel mutations generated by gRNA 4 and its corresponding target sequence and cut site 78

Figure 3.18 Illustration of the most common deletions generated by gRNA 6 and its corresponding target sequence and cut site 79

Figure 3.19 Miseq sequencing data of amplicon 3 for wt NIKS cells 79

Figure 3.20 Graphic presentation of the most common deletions in cell population transfected with gRNA 6 80

Figure 3.21 The consequence of single nucleotide deletion within *SYCP2* gene 80

Abbreviations

AE	Axial element
ARLD	Armadillo-repeat-like-domain
ATAD2	ATPase family AAA domain containing 2
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3-related
Cas9	CRISPR-associated nuclease 9
CE	Central element
CENP	Centromere protein
CIN3	Grade 3 cervical intraepithelial neoplasia
CRISPR	Clustered regularly interspaced short palindromic repeats
CSCC	Cervical squamous cell carcinoma
CTA	Cancer testis antigen
DMEM	Dulbecco's Modified Eagle Medium
DSB	Double strand break
dsDNA	Double stranded DNA
dsRNA	Double stranded RNA
GFP	Green fluorescence protein
gRNA	Guide RNA
HDR	Homology-directed repair
HNSCC	Head and neck squamous cell carcinoma
HORMAD	HORMA domain-containing proteins
HPV	Human papillomavirus
HR	Homologous recombination

ICC	Invasive cervical cancer
LCR	Long control region
LE	Lateral elements
MHC	Major histocompatibility complex
NHEJ	Non-homologous end-joining
NIKS	Normal immortalised keratinocytes
NLS	Nuclear localisation signal
NPI	Normalised percent inhibition
OPSCC	Oropharyngeal squamous cell carcinomas
ORF	Open reading frame
PAMs	Protospacer adjacent motifs
Rb	Retinoblastoma
RISC	RNA-induced silencing complex
RNAi	RNA interference
RT-qPCR	Real time quantitative polymerase chain reaction
SC	Synaptonemal complex
SCCHN	Squamous cell carcinoma of the head and neck
siRNA	Small interfering RNA
STAG3	Stromal antigen 3
SYCE	Synaptonemal complex central element protein
SYCP2	Synaptonemal complex protein 2
TBP	TATA-binding protein
TEX12	Testis expressed protein 12
TF	Transverse filament
URR	Upstream regulatory region

Cindy Dong

Abstract

Virtually all cervical cancer cases worldwide are caused by oncogenic (high risk) HPV genotypes, with 70% attributable to the two dominant types, HPV16 and 18. The HPV replicative cycle relies on the capability of keratinocytes to undergo differentiation in the stratified epithelia. However, persistent mucosal high risk HPV infections can result in abortion from the productive replication cycle and deregulated viral oncogene (E6 and E7) expression, driving carcinogenesis. A group of genes known as CTAs normally exhibit testis-specific expression in healthy individuals. The immune privileged characteristic of testis has allowed the identification of cancer testis antigens (CTAs) with aberrant expression in tumours as potential biomarkers for immunotherapeutic approach to cancer treatment. A component of the meiotic synaptonemal complex, SYCP2 is strongly upregulated in HPV+ compared to HPV- cancers and normal epithelia. In this project, we aim to measure SYCP2 expression in HPV+ cancer cell lines and determine the effect of gene knockdown using siRNA or gene knockout using CRISPR-Cas9. An SYCP2 qPCR assay was established and optimised and SYCP2 overexpression was confirmed in HPV16+ cell lines but intriguingly not in HeLa (HPV18+), suggesting some key cellular differences between these two genotypes. Moreover, cell viability did not seem to be affected by SYCP2 depletion, although complete gene silencing was not achieved. We also used next-generation amplicon sequencing to assess CRISPR gene editing efficiency of SYCP2, indicating the need for improved enrichment of transfected cells using fluorescence-activated cell sorting (FACS). This study has obtained data that complies with past studies on the upregulation of SYCP2 expression which will allow future work to elucidate the functional consequences of SYCP2 expression in HPV+ cancers.

1 Introduction

1.1 Human papillomavirus

1.1.1 Structure and genome

The *Papillomaviridae* is a family of double-stranded DNA viruses that demonstrate strong epitheliotropism and are known to infect many mammalian species and others such as birds and reptiles (Bravo, de Sanjosé and Gottschling 2010; Van Doorslaer *et al.* 2018). Human papillomavirus (HPV) comprises a large group of diverse viruses that exclusively infect the keratinocytes of mucosal and cutaneous stratified squamous epithelia in humans (Orav *et al.* 2015). Regions such as the genitals, anus and oropharynx are usual targets of the mucosal targeting HPVs (Tuominen *et al.* 2018). The majority of humans would have been colonised by several HPVs at some point in their lifetime, causing asymptomatic infections which are generally eliminated by the immune system (Bravo and Felez-Sanchez 2015). HPV triggers host cell proliferation as part of its infectious cycle, and thus HPV infection may lead to benign clinical manifestations such as skin and genital warts (papilloma). More rarely, HPV can become integrated into the host genome. This disrupts the regulation of both virus and host genes and can thus lead to cancer development through uncontrolled expression of viral proto-oncogenes (Handisurya, Schellenbacher and Kirnbauer 2009).

To date, more than 200 HPV genotypes are known (Schiffman *et al.* 2016) and are categorised into five phylogenetic genera based on disease association, replicative cycle features and DNA sequencing data, designated as alpha, beta, gamma, mu and nu (Doorbar *et al.* 2012; de Villiers 2013), as shown in Figure 1.1. The alpha types mostly

infect mucosal epithelia, whereas the beta-HPVs generally infect cutaneous epithelia (Tomaić 2016).

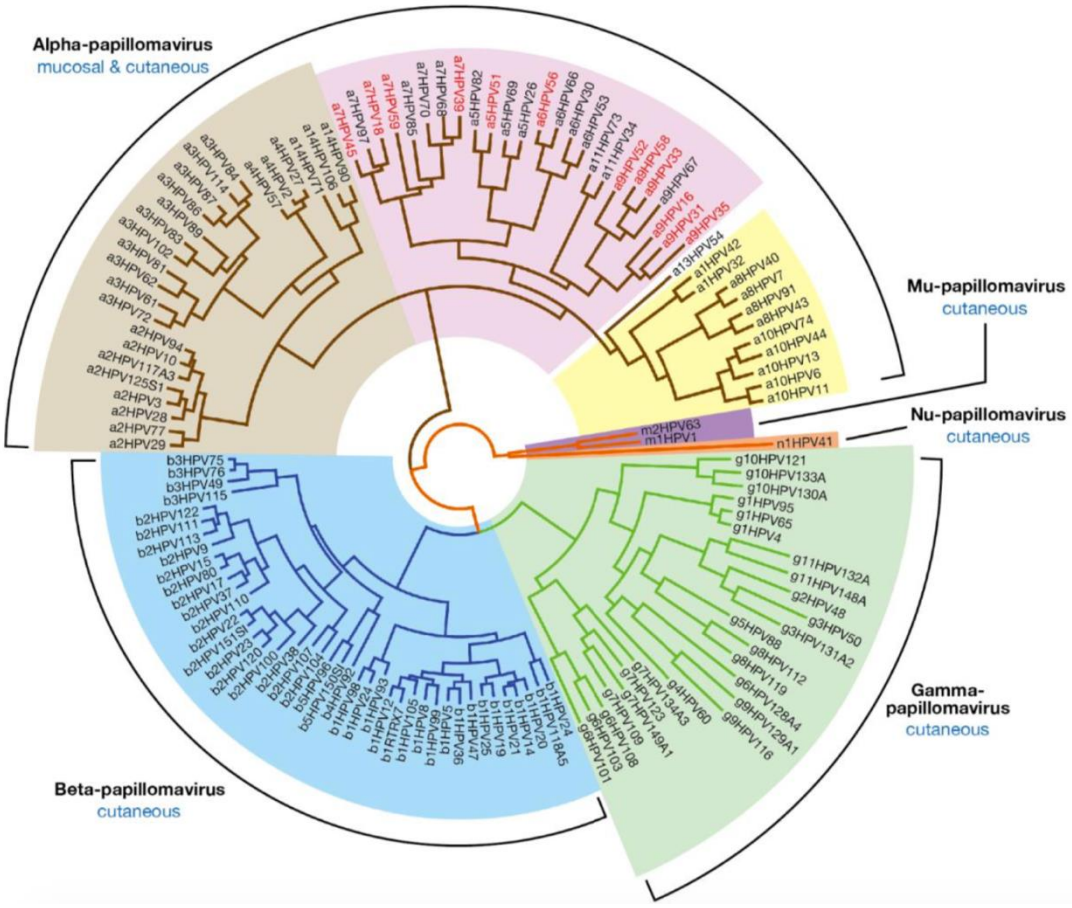


Figure 1.1 The phylogeny of HPV

Illustration of the five HPV genera and the evolutionary relationship between different HPV genotypes. (Doorbar *et al.* 2012)

HPVs are non-enveloped viruses that are about 60nm in diameter, with an 8-kilobase, circular double-stranded DNA (dsDNA) genome contained within an icosahedral protein capsid comprised of 72 pentameric capsomers (Sapp 2013; OMS 2007). All HPV types include approximately eight open reading frames (ORFs) in their genome, which can be divided into three main regions: early, late and a non-coding segment known as the long control region (LCR) or the upstream regulatory region (URR) (Zheng and Baker 2006; OMS 2007). The well-characterised structure of HPV16 genome is shown in Figure 1.2. Individual HPV types are distinguished from others by having at least 10% difference in the nucleotide sequence of L1 ORF (Liu *et al.* 2017). Those with differences between 1-

10% (within the same HPV types) are known as HPV variant lineages (Harari, Chen and Burk 2014).

The viral genome mostly encodes six regulatory proteins from the early region: E1, E2, E4, E5, E6 and E7. Plus, two structural proteins from the late region: L1 and L2, the major and minor capsid proteins, respectively (Graham 2010; Depuydt *et al.* 2016). The LCR region covers approximately 10% of the HPV genome, containing the origin of replication and many DNA recognition sites for both host and viral transcription factors. These are essential for regulating gene transcription from the early and late promoters (Zheng and Baker 2006). The genome also comprises of a keratinocyte-specific enhancer region adjacent to the early gene promoter (Harari, Chen and Burk 2014).

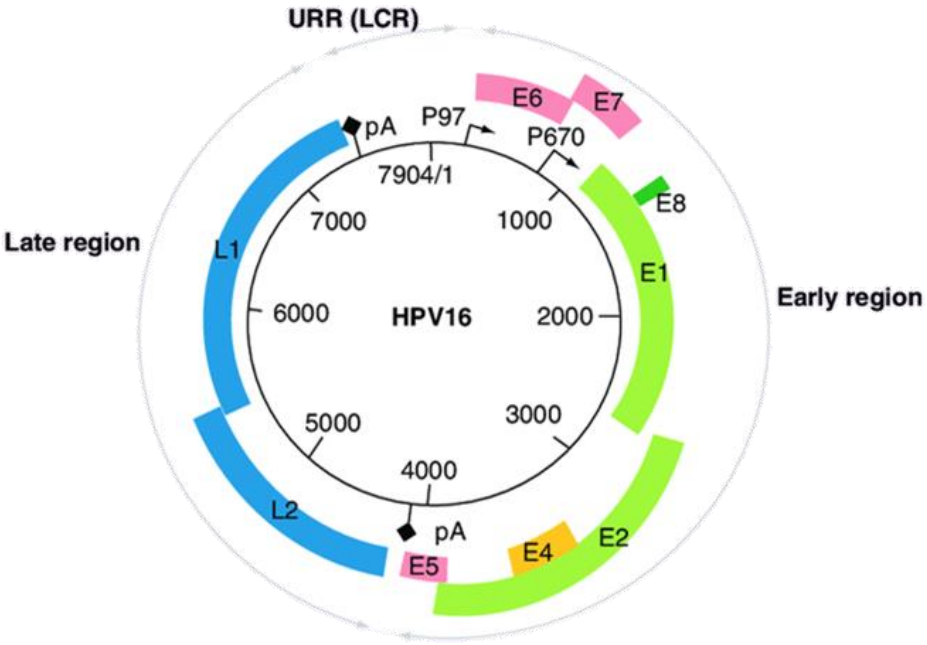


Figure 1.2 HPV16 genome structure

Illustration of HPV16 genome with 7904 bp, comprising of three main segments. The early region that encodes seven ORFs E1, E2, E4, E5, E6, E7 and E8. The late region encodes two ORFs, L1 and L2. The LCR is a non-coding region. The ORFs in blue expresses capsid proteins. The pink ORFs encodes HPV16 oncoproteins and the green ORFs expressions proteins vital for HPV replication and transcription. (Nakahara and Kiyono 2016)

Both E1 and E2 protein participate in viral genome replication where E1 oligomers act as an ATP-dependent helicase to separate the dsDNA and transcription factors are encoded by the E2 ORF. Heterodimers formed by E1 and E2 join the cellular replication machinery at the origin of replication, from which bidirectional DNA synthesis is triggered (Reinson *et al.* 2013; Harari, Chen and Burk 2014). Due to messenger RNA (mRNA) splicing, E4 protein exists as an E1^E4 fusion protein and is abundantly expressed in productive lesions, especially in cells within the upper epithelial layer, suggesting a role in genome amplification to assist virus synthesis. It has also been shown to function in viral release both *in vivo* and *in vitro* as its N-terminal leucine-rich motif (LLXLL) associate with and reorganises the cytokeratin network (Doorbar 2013; Egawa *et al.* 2017). Furthermore, in some HPV types (HPV18, 31 and 11) analysed in human primary foreskin keratinocytes where there is disruption to the keratin binding motif, genome amplification is impaired. However, the disruption does not always interfere with the viral replicative cycle across all HPV types. (Nakahara *et al.* 2005; Egawa *et al.* 2017).

E5 proteins are not expressed by all HPV genera. For example, beta-, gamma- and mu-HPVs do not contain the E5 ORF but alpha-HPVs do (Di Domenico *et al.* 2009). This indicates that it is not important for the completion of the HPV replication cycle. Instead, it may provide some form of advantage during infection. E5 proteins are found in the suprabasal layer and functions in apoptosis inhibition and upregulation of transcription factors to maintain active cell division (Müller *et al.* 2015). Also, it can reduce the expression of major histocompatibility complex (MHC) I on the surface of infected basal cells, inhibiting antigen presentation to T-cells. Thus, E5 may help to evade the immune surveillance (Venuti *et al.* 2011).

Expression of E6 and E7 is essential for the HPV replicative cycle. Early studies have shown increased mRNA level in the lower to mid-upper epithelial layers (Stoler *et al.* 1992) but reduction in E6 and E7 biomarker expression was seen in the upper epithelial layers (Middleton *et al.* 2003). These results indicate that E6 and E7 proteins are required in the early stages of HPV replication. E6 has been shown to play a role in maintaining episomal genome (Park and Androphy 2002).

Upon E6 expression, the ubiquitin-proteasome system is hijacked to facilitate rapid p53 degradation by forming a trimeric complex consisting of p53, E6 and E6AP (an E3 ubiquitin ligase). The loss of p53 function is detrimental for the cells, it results in lack of apoptosis regulation, increased genomic instability, disruption in the cellular response to DNA damage and other cellular functions (Ozaki and Nakagawara 2011). E6 also interferes with many other biological functions such as the cell cycle, G1 checkpoint activity is lost very early on in transformed cells expressing E6 (Yim and Park 2005).

Non-phosphorylated retinoblastoma (RB) behaves as a transcriptional repressor by binding at promotor regions containing E2F (transcriptional factor) sites. E7 protein can bind to the 'pocket domains' of non-phosphorylated RB and cause degradation through the ubiquitin-proteasome pathway. This result in the release of the transcriptionally active form of E2F as the association between RB and E2F is lost. Consequently, transcriptions induced by E2F are upregulated including key components of the cell cycle (CDK2, cyclins A and E). Causing stimulation of the checkpoint between G_1 and S phase in basal cells that undergoes differentiation. Therefore, it allows the expansion of cells undergoing DNA replication in the basal layer (Yim and Park 2005; Tomaić 2016; Graham 2017). Later in the infection, E7 protein also prompt S phase re-entry of differentiated

keratinocytes. This is accomplished by destabilising a RB related pocket protein, p130 which is known to prevent S phase re-entry (Banerjee *et al.* 2011).

1.1.2 The infectious cycle of HPV

Upon micro-trauma of the stratified epithelium, infectious HPV particles come into contact with the basal keratinocytes where they release the HPV genome into the host cells. This genome is then transported to the nucleus for transcription and replication. The differentiation stages of the multi epidermal layers is relied on by the process of HPV replication cycle (Reinson *et al.* 2015).

In vivo, HPV infections are established through initial binding between L1 proteins and the heparin sulphate proteoglycans (such as Syndecan-1) on the extracellular basement membrane. This induces a cyclophilin B-induced conformational change where it exposes the N-terminus of L2 proteins on the surface, allowing removal of one of its nuclear localisation signal (NLS) domains via cleavage by cellular furin protease (Graham 2017). The consensus furin cleavage motif (R-X-K/R-R) is found conserved across numerous papillomavirus types (Buck, Day and Trus 2008; Day, Lowy and Schiller 2008; Wang and Roden 2013). Further conformational change of the capsid occurs which may lower the affinity of L1 for heparin sulphate proteoglycans and expose particular binding sites on L1, allowing interactions with yet to be identified secondary receptors on the surface of basal keratinocytes. This prompts internalisation of the viral particles via tetraspanins-mediated endocytosis (Richards *et al.* 2006; Horvath *et al.* 2010). In recent studies, there have been some indications that epidermal growth factor receptors (Surviladze, Dziduszko and Ozbun 2012), integrins (α6 integrin) and laminins (Culp *et al.*

2006) are potential candidates as secondary receptors that facilitate the entry of HPVs. Interestingly, studies of cells with furin deficiency and presence of furin inhibitors *in vivo* and *in vitro* demonstrated low level of infection activity across various papillomavirus types (Wang and Roden 2013). Thus, indicating the importance of truncated L2 in inducing effective infection.

Upon entry, HPVs are transported in vesicles through the early endosomes, late endosomes/lysosomes, trans-Golgi-network and the endoplasmic reticulum before they are directed to the nucleus where they enter the nuclear pores after nuclear envelope breakdown in mitosis (Aydin *et al.* 2014; Aksoy, Gottschalk and Meneses 2017). Viral capsids are degraded in the endolysosomal system and L2 proteins are known to mediate trafficking of the viral genome to the trans-Golgi-network (Aksoy, Gottschalk and Meneses 2017). Once the genome is in the nucleus, it has been shown to reside at the PML oncogenic domains (PODs) where viral gene transcription are initiated (Broniarczyk *et al.* 2018).

Firstly, HPV undergoes an initial amplification of its genome to achieve a viral copy number of approximately 50 to 100 copies per keratinocyte (Stanley 2012). Only the early transcripts are transcribed at this point, primarily for their role in genome replication. Once the infection is established, the viral copy number is maintained at a low level in the basal layer and transcription also occurs at a low rate (Bienkowska-Haba *et al.* 2018).

As the infected cell divide into two daughter cells, the one that has initiated differentiation (known as transit-amplifying cells) migrates into the suprabasal layer

whilst carrying the viral genome. It was confirmed in a study (Oldak *et al.* 2004) that E2-mediated integrin suppression was required for the detachment of infected keratinocytes from the basement membrane. In particular, high concentrations of β4-α6 integrin dimers are found in the basement membrane (Oldak *et al.* 2004). As migration occurs, HPVs enters the final stage of its replication cycle, the vegetative amplification phase. It induces G2 arrest in the proliferating keratinocytes (Wang *et al.* 2009) and uses them to extensively replicate its viral genome. This is an unusual strategy for virus replication. It could be the case that HPV is unable to hijack the replication machinery whilst the host undergoes DNA replication. However, by inducing G2 arrest, it would provide a chance for HPV to perform rapid and efficient amplification without any competition.

In the upper epithelial layers, the level of viral proteins increases significantly, leading to amplification of thousands of genome copies per cell (Kajitani *et al.* 2012). In the terminally differentiated cells, capsid proteins (L1 and L2) are synthesised for the assembly of virions. The viral genomes are packaged into the progenitor virions and are released from the keratinocytes in the cornified epithelial layer (mostly the upper layer) (Kajitani *et al.* 2012).

The DNA damage response pathway plays a major role in genome amplification and the expression of late genes. Ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3-related (ATR) kinases are two important sensors of DNA damage within cells and regulators of the repair mechanism (Graham 2017). Early in the replication cycle, E1-mediated ATM activation may contribute to the maintenance of HPV genome copy number in the basal keratinocytes. Furthermore, numerous DNA damage repair proteins

were contained within the replication factories that are formed when E1 and E2 come together, facilitating the viral genome replication (Blackford and Jackson 2017). E7 has been demonstrated to activate the ATM pathway required for the vegetative amplification phase in differentiated epithelial cells (Banerjee *et al.* 2011).

HPV is capable of sustaining an infection for a prolonged period by retaining its episomal genome within the epithelial basal layer whilst expressing limited viral components. One of the ways to evade immune detection is through the expression of E2 protein, along with cellular proteins such as SMCX (a methylase), EP400 (a component of the TIP60 histone acetyltransferase complex) and Brd4 (an E2-binding bromodomain protein). Together, they transcriptionally repress the early promoter gene by altering the chromatin conformation (Smith *et al.* 2014; Graham 2017) which reduces the expression level of E6 and E7.

In HPV 1, 8, 16 and 31, E8^E2 fusion protein interacts with the corepressor complexes which act as a negative regulator of viral replication and transcription. Notably, the protein also plays a critical role in limiting the productive replicative cycle of HPV16. Therefore, the expression level of E8^E2 protein may determine if the infection progresses or becomes latent (Dreer, van de Poel and Stubenrauch 2017).

HPVs can also compromise the innate immune system in infected keratinocytes by suppressing synthesis and signalling of interferons (IFNs) such as IFN-α and IFN-β. IFNs are essential for the control of viral infection. Cell deaths are not observed from the replication cycle and the assembly of virions which leads to lack of inflammation. Therefore, the keratinocytes release a little or no pro-inflammatory cytokines,

suggesting that the antigen-presenting cells such as Langerhans cells and stromal dendritic cells are not activated and migrate to the site of infection. As a result, the virus essentially delays the activation of adaptive immunity and remain undetectable to the host defence system (Stanley 2012).

1.1.3 HPV and cancer

1.1.3.1 HPV prevalence

Globally, HPV is the most frequent sexually transmitted infection and it has been proposed that about 75% of men and women who are sexually active will be affected by genital HPV infection at some stage in their life (Nejo, Olaleye and Odaibo 2018). Approximately 11-12% of women worldwide with normal cytology were tested positive for HPV in a meta-analysis of cervical HPV prevalence (Bruni *et al.* 2010) involving 194 studies and test subject of over one million women. However, the prevalence varied considerably between different regions of the world. It is significantly higher in Eastern Europe (21.4%), Latin-America (16%) and rises to above 30% for Eastern Africa and the Caribbean. Regions with the lowest rate include Northern America (4.7%) and Western Asia (1.7%) (Formana *et al.* 2012).

Another worldwide meta-analysis (Guan *et al.* 2012) including 423 PCR-based studies, compared 266,611 women with normal cytology and 106,625 with cervical abnormalities. The results coincide with the former meta-analysis where 12% of women with normal cytology were HPV-positive. The study also showed a strong increase of prevalence in HPV positivity rate as lesion severity rises, reaching the highest of 93% in women with grade 3 cervical intraepithelial neoplasia (CIN3). 89% of women with

invasive cervical cancer (ICC) were tested positive for HPV, which is the second-highest rate among all cervical disease grades (Formana *et al.* 2012). There was also a significant increase in HPV16 detection among HPV-positive women as cervical lesion severity increases, with the highest of 63% in women with ICC.

1.1.3.2 HPV-related cancers

The majority of cutaneous and mucosal HPV types, including HPV 6 and 11 within the alpha genus are classified as low-risk, causing respiratory papillomatosis and genital warts which are not life-threatening and rarely cause neoplasia (Danielewski *et al.* 2013). They are usually cleared up by the host immune system without any intervention within two years (Salman *et al.* 2017).

On the contrary, high-risk HPV types including HPV16, 18 and 45 are well-established cancer-causing agents in areas including the cervix, anus, penis, vulva, vagina, and the oropharynx. HPV16 and 18, the two most carcinogenic HPV genotypes are accountable for about 70% of cervical cancer worldwide, with HPV16 being the predominant causative agent (Pirog *et al.* 2014). HPV16 is also the most prevalent type in HPV+ head and neck cancers whereas other high-risk types are hardly detected (Tomaić 2016).

Cervical cancer is the most common HPV-induced cancer and a massive health burden globally. It was ranked fourth for incidence and mortality for women in a global cancer statistic report (GLOBOCAN) produced by the International Agency for Research on Cancer in 2018 (Bray *et al.* 2018). Notably, it was the leading cause of cancer-induced mortality in 42 countries out of 185. It has been estimated that by 2040, the number of new cervical cancer incidences would rise to over 770,000 per annum, compared to

569,000 in 2018. Mortality rate would also increase to over 450,000 per annum globally, compared to 311,000 in 2018 (Bray *et al.* 2018). Notably, the less developed regions contribute a major portion to these statistics. The mortality age-standardised rate per 100,000 females in low/medium human development index (HDI) regions were 18.2 in comparison to 4.1 in high/very high HDI regions (Bray *et al.* 2018).

1.1.3.3 HPV genome integration

Genomic integration represents an abortion of the HPV replication cycle; it is frequently associated with truncations or deletions of viral genes (Pinidis *et al.* 2016) and once integrated into the host genome, the viral genome can no longer be amplified or packaged and transmitted to a new host. Thus, it is non-productive for the virus. HPV integration is detected in premalignant cervical lesions, but is more frequently detected in invasive carcinoma (McBride and Warburton 2017). Integration is thought to promote oncogenesis by deregulating the expression of E6 and E7 genes to enhance proliferation, disruption of cell cycle checkpoints and triggers progressive genomic instability (McBride and Warburton 2017).

A model suggested that cervical inflammation could promote HPV genome integration in cervical carcinoma (Williams *et al.* 2011). It is thought that the generation of inflammation-induced reactive oxygen and nitrogen species could result in DSBs in both host and viral genome. Consequently, allowing integration to occur and promote carcinogenesis.

1.1.3.4 HPV oncoproteins

E6 and E7 are the two predominant proteins that contribute to the oncogenesis of high-risk HPV type-induced cancers. They are transcribed as a single bicistronic transcript due to being located in the same ORF (Jiang and Yue 2013). Their contribution toward carcinogenesis includes immortalisation of proliferating cells and cell cycle disruption through degradation or inactivation of essential cell cycle regulatory components including the tumour suppressor proteins p53 and Rb (Vishnoi *et al.* 2016; Jing *et al.* 2014). High levels of E6 and E7 expression have been demonstrated to induce polyploidy by disrupting the G_2-M-phase transition in primary human keratinocytes (Patel *et al.* 2004).

Before host genome integration, low levels of E6 and E7 are found in the basal cells (Nees *et al.* 2000), mostly due to the positioning of E2 binding sites within the E6/E7 promoter region. Hence, it was proposed that the breakpoint of HPV episome occurs at the *E1* and *E2* gene which corresponds to the transcription repression effect of E2 on E6 and E7. Therefore, the loss of E2 expression induced by integration directly result in upregulation of E6 and E7 as cancer progresses (Francis, Schmid and Howley 2000). However, this is not always the case. There are also many other HPV-associated cancer (particularly HNSCC), in which the HPV appears to remain episomal.

Expression of E6 and E7 oncoproteins provide a suitable setting for DNA replication in differentiating epithelium which drives oncogenesis. However, tumour development is merely a side effect of HPV infection. The ability of host cells to replicate is an advantage for HPV to perform genome amplification and sustaining infections.

1.2 Cancer testis antigens

1.2.1 CTA expression

CTA comprise a group of genes that normally demonstrate exclusive expression in the germ line tissues such as testis, placenta and ovary of healthy individuals. However, low level expression of CTAs has also been found in some somatic tissues or immune-privileged areas. CTAs that are expressed in testis and no more than two other tissue types are classified as testis selective. Some CTAs are termed testis brain restricted when it is expressed in testis and the central nervous system. CTAs that are limited to the testis are testis restricted (Whitehurst 2014).

To date, there are over 270 CTA genes (from CTDatabase) that are divided into two main categories, over 100 of which are cancer testis-X (CT-X) antigens that are encoded on the X chromosome and another 100 genes are located on autosomes, producing the non-X CTAs (Bode *et al.* 2014; da Silva *et al.* 2017), with the remainder found on the Y chromosome (Dobrynin *et al.* 2013). X-CTA genes are predominantly expressed on the spermatogonia in normal testes. Whereas, non-X CTAs are expressed on spermatocytes that are in the late stages of germ-cell differentiation (Fratta *et al.* 2011).

The discovery of CTAs and their increased expression in a broad spectrum of cancers were significant in tumour immunotherapy. The absence of human MHC-I molecules in germ cells implies that they cannot undergo antigen presentation to T cells (Bruggeman *et al.* 2018). Hence, CTA expressed in tumour cells can be regarded as neoantigens, thus triggering humoral and cell-mediated immune responses which are strictly cancer-specific (Gjerstorff, Andersen and Ditzel 2015). This has identified CTAs as promising

biomarker targets for immunotherapeutic approaches to cancer treatments such as adoptive T-cell transfer with chimeric T cell receptor and CTA vaccine therapy (Fratta *et al.* 2011; da Silva *et al.* 2017).

The aberrant expression of CTAs in tumours is an intriguing concept, the underlying mechanisms responsible for the upregulation is yet to be fully understood. Whereas in normal and neoplastic cells, it is suggested that epigenetic events play a major role in CTA expression. Histone acetylation and DNA methylation are regarded as two of the most well-characterised factors (Fratta *et al.* 2011). Most of CTAs in normal somatic cells contain methylated CpG islands in their promoter region which repress expression. However, they are activated by demethylation during spermatogenesis in germ line cells. Another epigenetic modification is deacetylation of histone, leading to chromatin compaction which supresses the activity of transcription factors and RNA polymerases. Thereby, acetylation and inhibition of deacetylation enzymes can stimulate expression of CTAs (Karpf 2006; Salmaninejad *et al.* 2016). Transfection experiments with reporter genes driven by methylated/unmethylated CTA promoters *in vitro* were the decisive evidence for association between hypomethylated CTA promoters and their elevated expression (Luca Sigalotti *et al.* 2002; L. Sigalotti *et al.* 2002).

1.2.2 Functional role of CTAs in cancer

In previous studies, several CTAs have been suggested to play a positive regulatory role in signalling pathways that are often hyperactive in cancerous cells. An example is ATPase Family AAA Domain Containing 2 (ATAD2) protein which behaves as a cofactor for MYC-mediated transcriptions. ATAD2 is activated by E2F in Rb-deficient cancers and

it has been proposed that the protein may promote tumourigenicity by enhancing the MYC-dependent transcription, leading to disruption of several key cellular processes such as cell growth, DNA damage response and genomic instability (Adhikary and Eilers 2005). Reduction of ATAD2 level in tumour cells impedes with cellular proliferation by preventing S-phase entry (Whitehurst 2014). Furthermore, its high expression level corresponds with poor prognosis in breast cancer patients (Ciró *et al.* 2009; Kalashnikova *et al.* 2010).

PRAME (melanoma antigen preferentially expressed in tumours) is another CTA that is overexpressed in human malignancies including melanomas, head and neck cancers and chronic leukaemia (Whitehurst 2014). A study identified PRAME as a dominant repressor of retinoic acid receptor signalling (Epping *et al.* 2005). It prevents intrinsic ligand (retinoic acid)-receptor interaction by occupying the receptor binding site. This signalling pathway is particularly important for proliferation arrest, differentiation and apoptosis. Thereby, disruption of this pathway is frequently observed in cancers and it was suggested that upregulated expression of PRAME provided growth or survival advantages for the tumour cells (Epping *et al.* 2005).

1.3 Synaptonemal complex protein 2

1.3.1 Meiotic function

Synaptonemal complex protein 2 (SYCP2) is a meiotic-specific protein of molecular mass of 175KDa and it has been observed to be overexpressed in HPV-related cancers. However, it is not officially categorised as a CTA like its family member SYCP1 despite their almost exclusive expression in the testis.

To date, not much information is known about this protein. However, it is believed to be involved in chromosome linkage through the synaptonemal complex (SC) and to drive meiotic prophase I. Prophases are generally divided into five substages, leptotene, zygotene, pachytene, diplotene and diakinesis. During leptotene, homologous chromosomes are paired up, SYCP2 and SYCP3 form a heterodimeric structure at the interchromatid domain called lateral elements (LEs). Before synapsis of homologues, LEs are referred to as axial elements (AEs) (Pelttari *et al.* 2001; West *et al.* 2019). In zygotene, the transverse filaments (TFs) comprises of SYCP1 acts as a linker between the two LEs and another component of the SC, known as central element (CE). CEs are primarily composed of proteins such as synaptonemal complex central element protein (SYCE) 1, 2, 3 and testis expressed protein 12 (TEX12) (Seo *et al.* 2016). Together, they form the zipperlike structure known as SC, shown in Figure 1.3. Complete assembly occurs in pachytene where synapsis is complete and in diplotene, the disassembly is initiated (Pelttari *et al.* 2001; Fraune *et al.* 2012). In all diploid organisms where SCs are found, they share this conserved tripartite structure which signifies the importance of SC in meiosis progression.

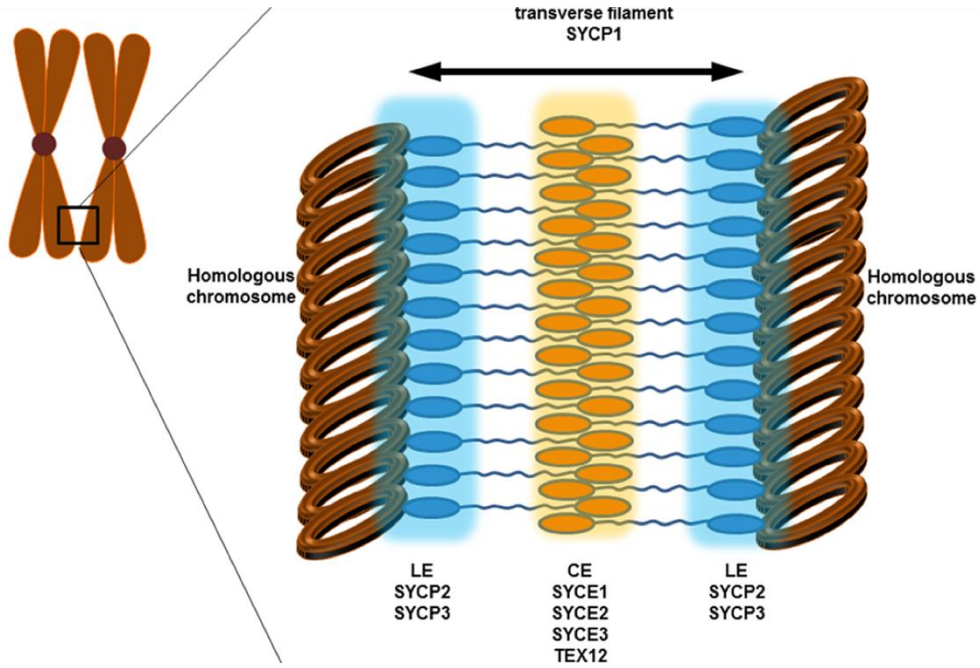


Figure 1.3 Structure of synaptonemal complex

Illustration of components that make up the synaptonemal complex. The transverse filament links the LE and CE together. SYCP2 and SYCP3 of the lateral elements interact with the homologous chromosomes. (Seo *et al.* 2016) LE = lateral element, CE = central element

A common feature of DNA binding proteins is Ser/Thr-Pro and Ser/Thr-Ser/Thr motifs, allowing binding with DNA at the minor groove in a non-sequence-specific manner. SYCP2 in mice contains two large clusters of these motifs near its C-terminus which are proposed to form a short coiled coil domain (Offenberg *et al.* 1998). This domain is conserved throughout evolution and is required for interaction with SYCP3 and chromatin. In mice *SYCP2*$^{-/-}$ spermatocytes with deleted coiled coil region, SYCP3 form large protein aggregates in the nucleus but incorporation into the SCs was not detected (Yang *et al.* 2006). This confirms that presence of SYCP2 is necessary for SYCP3 localisation to the axial chromosome core. Furthermore, lack of SYCP2 also impairs LE formation. Silver nitrate is used as a stain to identify LEs and paired LEs. Electron microscopy reveals no silver-stained LEs in *SYCP2*$^{-/-}$ spermatocytes, structures that resemble SC with alignment between chromatins and CE-like structures were formed instead. However, the structure does lack the electron dense LEs that are generally

observed in wild-type spermatocytes (Yang *et al.* 2006). In *SYCP3*-deficient meiotic cells, no silver-stained LEs were detected, illustrating that SYCP3 is critical for SYCP2 recruitment, LE assembly and potentially plays a role in anchoring LEs to meiotic chromatin (Pelttari *et al.* 2001).

The formation of SC is indispensable for the effective segregation of homologous chromosomes during the first meiotic division. In male mice, depletion of SYCP2 or SYCP3 has been found to result in severe synaptic defects and apoptosis of spermatocytes, resulting in infertility (Yang *et al.* 2006). However, the effect is fairly mild in *SYCP3*-deficient females, the females become sub-fertile with reduced litter size, caused by increasing number of aneuploidies (Yuan *et al.* 2002; Yang *et al.* 2006). This indicates that mice may exhibit sexually dimorphic response towards meiotic disruption caused by the loss of LE proteins (Alves *et al.* 2017). Association between defective SC formation and non-lethal aneuploidy such as Down syndrome have been demonstrated (Seo *et al.* 2016).

In mammalian meiotic cells, the sister chromatids must remain attached during first round of meiotic division. This is mainly achieved by a cohesin complex comprised of meiotic-specific proteins and those that are also present in mitotic cells (Kouznetsova *et al.* 2005). Additional functions of LEs were proposed in establishing and/or maintenance of cohesion between sister chromatid. Stromal antigen 3 (STAG3) is a component of the cohesin complex and was observed to colocalise with SYCP3 in diplotene spermatocytes and pachytene/late diplotene oocytes using immunostaining (Prieto *et al.* 2001). Interestingly, STAG3, SYCP3 and SYCP2 all seem to disappear within the first week of postnatal oocyte development. Additionally, in early dictyotene oocytes, short STAG3-

positive filaments were discovered almost only at sites where SYCP3 were still detected on chromatin (Kouznetsova *et al.* 2005). Dictyotene is unique to oogenesis, causing complications when making direct comparison to male germ cells. Although, analysis of cells at prometaphase I during spermatogenesis also shows a fragmented cohesin core (Prieto *et al.* 2001). The observations indicate a possible mechanistic link between the loss of the two LE proteins and the cohesin core.

It is proposed that SYCP3 may have the ability to modulate integrity of cohesin core. *SYCP3*-deleted spermatocytes have shown cohesin core formation in zygotene and pachytene (Pelttari *et al.* 2001). However, as cells advanced into the diplotene stage, the STAG3 cohesin cores became highly fragmentated. This confirms the need of LE proteins for desynapsed STAG3 cohesin core integrity, as well as preventing premature disassembly of the cohesin core complex (Kouznetsova *et al.* 2005). Together with SC, both complexes contribute to the organisation of the chromosome axis during meiotic prophase I and ensure proper chromosome segregation.

The N-terminus of mouse SYCP2 seems to associate with the centromere region during meiosis I. This is achieved by an ARLD (armadillo-repeat-like-domain) subdomain interacting with two components of the centromere, CENP (centromere protein) J and CENP F (Feng *et al.* 2017). Given the interaction between the C terminus of SYCP2 and SYCP3/SYCP1, along with centromere association, it is rational to postulate that SYCP2 may behave as a bridge between SC and the centromere. This idea is consistent with the lack of SYCP1 immunofluorescence and gradual disappearance of SYCP3 observation in the late diplotene stage (Feng *et al.* 2017).

HORMA domain-containing protein (HORMAD) 2 is a CTA that act as an important regulator of double strand breaks (DSBs) and crossover formation in meiosis (West *et al.* 2019). It is proposed that the chromosome axial core proteins recruit HORMADs to the axis via their HORMA domain-binding closure motifs, which are later removed by AAA+ ATPase, thyroid hormone receptor interactor 13 (TRIP13) as germ cells enter the pachytene stage. This prevents further DSBs and allowing recombination of homologues (Wojtasz *et al.* 2009). It has been shown that SYCP2 interacts with HORMAD2, suggesting a possible a role of SYCP2 in HORMAD2 recruitment.

1.3.2 Overexpression of SYCP2 in HPV-positive cancers

Several gene expression profiling studies of HPV-positive cancers have shown upregulation of *SYCP2*, along with other characterised CTAs (Slebos *et al.* 2006; Martinez *et al.* 2007; Masterson *et al.* 2015). In one of the studies (Schlecht *et al.* 2007), 29% of 42 head and neck squamous cell carcinoma (HNSCC) patients were tested positive for HPV16. When focused on never-smokers, out of the 123 genes that were identified with higher expression than HPV16-negative primary HNSCC tumours, SYCP2 expression had the highest fold change of over 120-fold (shown as the absolute difference in expression ratios in inverse log scale).

A transcriptome analysis study (Masterson *et al.* 2015) carried out with a cohort of HPV-associated oropharyngeal squamous cell carcinomas (OPSCC) patients identified *SYCP2* as one of the most upregulated genes in premalignant (carcinoma *in situ*) tissue, validated using real time quantitative polymerase chain reaction (RT-qPCR). A significant increase relative to baseline expression was observed from fresh frozen HPV+ tissue

samples (log2 fold change of 1.8; P<0.01; 95 confidence interval [CI], 1.0-3.2). SYCP2 was included by The Wisconsin Alumni Group Foundation as one of the three potential biomarkers for OPSCC and precancerous lesions, along with STAG3 and testicular cell adhesion molecule 1, all of which exhibit testis-specific expression.

SYCP2 was also detected as one of the upregulated genes in a genome-wide expression profiling study (Pyeon *et al.* 2007), including 56 head and neck and cervical cancers where 38% of head and neck cancers and approximately half of cervical cancers were HPV16-postive. Further analysis of *SYCP2* in normal immortalised keratinocytes (NIKS) with HPV16 revealed an expression increase of >15-fold, in relative to HPV-negative NIKS. Knockdown of HPV16 E7 protein induced a significant decrease in SYCP2 expression relative to HPV16-postive NIKS. Interestingly, it did not result in the same expression level as HPV-negative NIKS, there was still approximately a 3-fold increase. The results suggested that SYCP2 expression was only partially induced by E7 (Pyeon *et al.* 2007).

Furthermore, fold changes of *SYCP2* expression (at the mRNA level) in primary cervical keratinocytes with E6 and E7 (delivered by recombinant retrovirus) were considerably higher than those with either one of the proteins expressed. This indicates a synergistic effect of E6 and E7 on SYCP2 expression. SYCP2 was also identified at the protein level in NIKS16 but not in NIKS, in parallel with the HPV16 E7 detection. However, it was not detected in CaSki, an HPV16+ cervical carcinoma cell line which had significant high level of HPV16 E7 protein (Pyeon *et al.* 2007). SYCP2 was also found to be among a set of genes that are consistently upregulated in HPV+ tumours, regardless of where they occur in the body (Chakravarthy *et al.* 2016).

1.4 Gene silencing/editing mechanism

1.4.1 RNA interference

RNA interference (RNAi) is an evolutionary conserved strategy used by a range of organisms to induce gene silencing. mRNAs are targeted by double stranded RNA (dsRNA) homologs, ultimately leading to their degradation. This phenomenon was first described by Fire and Mello in their Nobel prize study on genetic interference by exogenous dsRNA in *C. elegans* (Fire *et al.* 1998).

RNAi is divided into two main stages, the initiation, and the effector phase (Figure 1.4). In the first stage, dsRNA is cleaved in the cytoplasm by a specialised ribonuclease III-like enzyme called Dicer, resulting in a 21-23 nucleotides long dsRNA molecule, known as small interfering RNA (siRNA) (Lam *et al.* 2015). In the effector phase, siRNAs activate a multiprotein complex known as the RNA-induced silencing complex (RISC). The sense strand of siRNA undergoes cleavage by endonuclease argonaute 2 (component of RISC) and is disassociated from the complex, leaving the antisense (guide) strand attached to the RISC. Thus, the activated RISC is guided to its target mRNA sequence for cleavage and subsequent degradation (Agrawal *et al.* 2003; Petrova, Zenkova and Chernolovskaya 2013). Binding only occurs with fully complementary mRNA, thereby allowing highly specific gene silencing.

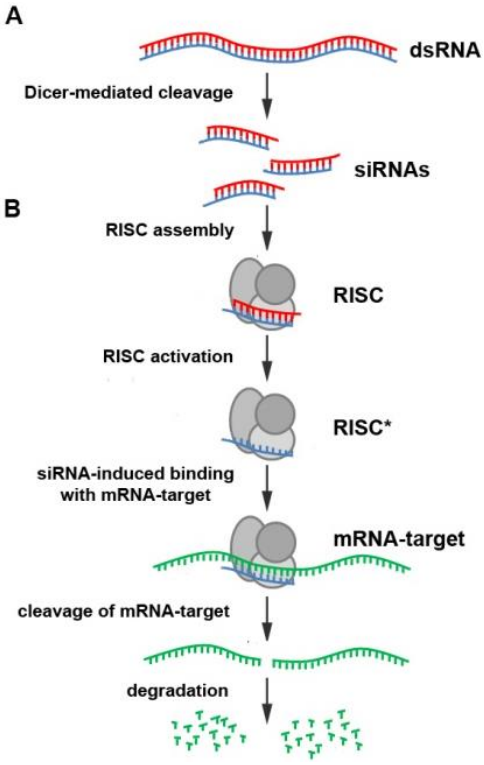


Figure 1.4 Mechanism of RNAi

RNAi comprises of two stages, (A) initiation and (B) effector phase. The dsRNA molecule is broken down into smaller fragments via Dicer-mediated cleavage, resulting in formation of siRNAs. Subsequent RISC assembly and activation allows the complex (RISK/siRNA guide strand) to the mRNA targeted by the siRNA guide strand. Consequently, the mRNA undergoes cleavage and becomes degraded. (Petrova, Zenkova and Chernolovskaya 2013) RISC = RNA-induced silencing complex

Due to its specificity, RNAi is regarded as a valuable tool in pharmacology and therapeutics. It has been introduced to patients with hereditary transthyretin amyloidosis where siRNA targets the liver (Setten, Rossi and Han 2019). Other uses of RNAi include gene expression analysis. In mammalian cells, RNAi can be stimulated by synthetically designed siRNAs that are 19-21 base pair in length. They essentially resemble the processed product by Dicer and can directly induce the effector phase. This feature allows RNAi to be utilised in targeted gene silencing for the study of expression and function.

1.4.2 CRISPR/Cas9 genome editing technology

Clustered regularly interspaced short palindromic repeats (CRISPR) along with CRISPR-associated nuclease 9 (Cas9) is a versatile and rapid genome manipulation system with high precision. Type II prokaryotic CRISPR system uses Cas9 derived from *Streptococcus pyogenes* and is the most commonly used system, requiring the presence of a sole Cas9 protein (Pellagatti *et al.* 2015). It is engineered and adapted from an ancient defence mechanism used by bacteria and archaea as part of their immune system. The mechanism allows the bacteria to defend against bacteriophages where they target and silence any detected foreign nucleic acid (Jinek *et al.* 2012).

The Type II system comprises of an endonuclease (Cas9) and a base-paired scaffold termed guide RNA (gRNA), combined from a sequence-specific targeting component (crRNA) and a trans activating CRISPR RNA (tracrRNA) that links crRNA and Cas9 together (Costa *et al.* 2004). To date, the CRISPR/Cas9 system is the most efficient genome editing technology that is adaptable to a wide range of applications including gene therapy (Costa *et al.* 2004).

In engineered CRISPR systems, the artificial gRNA is designed to contain a region of about 20 nucleotides at its 5'-end that is complimentary to the target DNA sequence. A conformational change of Cas9 protein is induced upon binding of gRNA, converting the Cas9 enzyme into its active form. The gRNA then directs the Cas9 to its corresponding protospacer adjacent motifs (PAMs) region (5'-NGG) that flanks the 3'-end of the target DNA sequence (Jinek *et al.* 2012). This allow the separation of DNA strands and subsequent binding between the complementary region of gRNA and the target strand. Resulting in cleavage of both target and non-target DNA strands by the conserved Cas9

histidine-asparagine-histidine (HNH) domain and RuvC domain, respectively (Jinek *et al.* 2012; Jinek *et al.* 2014).

Subsequently, a blunt end DSB was generated and it can be repaired by either of the two pathways (Figure 1.5): non-homologous end-joining (NHEJ) or homology-directed repair (HDR). Without the presence of an exogenous homology repair template, DSB can be repaired though the error-prone NHEJ pathway, allowing random insertion/deletion mutations (indels) or replacement (Xiao, Guo and Chen 2019). Alternatively, a donor template can be provided which will allow precise gene editing that is essentially error-free. Therefore, HDR can be used to induce a specific change within the target genome site (Pickar-Oliver and Gersbach 2019).

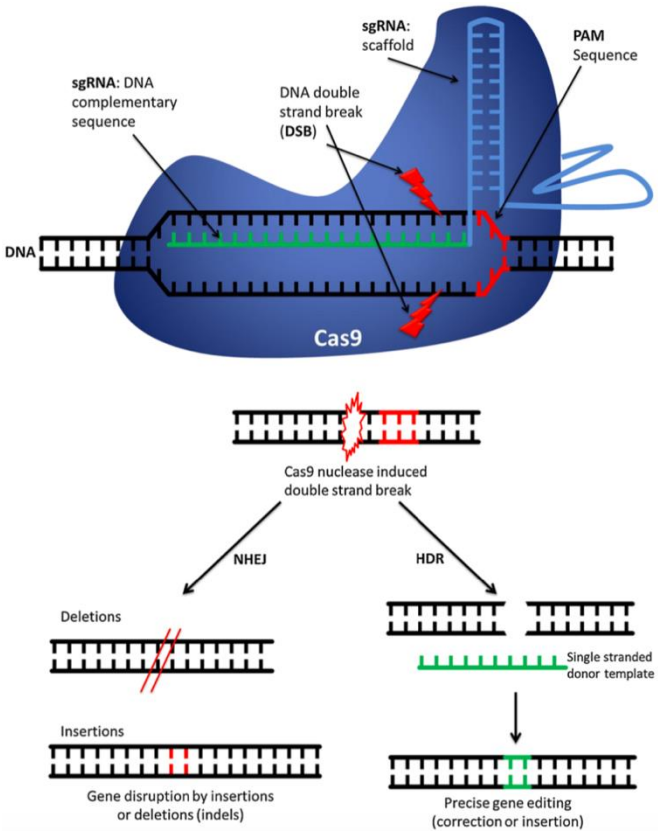


Figure 1.5 Mechanism of CRISPR-Cas9 gene editing technology

The CRISPR system composes of a Cas9 nuclease with a sgRNA complementary to the target sequence. Upon recognition, Cas9 induces DSB which can be repaired by either NHEJ or the HDR pathway. (Pellagatti *et al.* 2015) sgRNA = single guide RNA, DSB = double strand break, NHEJ = non-homologous end-joining, HDR = homology-directed repair

Due to its simplicity and flexibility, the CRISPR-Cas9 system has been widely exploited in numerous gene function studies, inducing both loss of function and gain of function mutations. Beyond genome editing, removal of the catalytic activity of Cas9 by inducing mutation in the RuvC and HNH nuclease domain has allowed it to act as a recognition complex, as it maintains its ability to target specific DNA sequence. This function enables fusion with transcription activators or repressors to stimulate site-specific gene regulation (Pickar-Oliver and Gersbach 2019).

1.5 Project outline

The ectopic expression of SYCP2 in HPV+ cancers and the lack of information on this protein has led to the interest of this project, hoping to unravel the reasons and concepts behind this unusual observation. It is in the interest of this study to shed light on these two questions:

1. What is causing SYCP2 expression in HPV+ cancers?
2. What is the relationship between SYCP2 and HPV+ cancer cells? Do they require SYCP2 for survival?

Upregulated SYCP2 expression in HPV+ cancer cells must be established as a starting point. This will be achieved by extracting RNA samples from seven cancer cells lines, five of which are HPV-positive and quantify expression level relative to housekeeping gene using qPCR. Three cell lines that demonstrate relatively high SYCP2 expression will undergo gene knockdown experiments to see the effect on cell viability using siRNA. Knockdown efficiency will be confirmed through qPCR.

In addition, CRISPR/Cas9 genome editing technology will be used in wild type NIKS cells to induce random indels mutations with five set of gRNAs targeting at different SYCP2 exon regions. The indels may cause frameshift mutations, leading to potential disruption in SYCP2 structure and function. Thereby allowing further examination of cellular responses to the loss of SYCP2. The plasmid contains *green fluorescence protein* (*GFP*) coding sequence, allowing the use of flow cytometry to sort and collect cells that were successfully transfected.

2 Materials and methods

2.1 Tissue Culture

2.1.1 Mammalian cancer cells

All the mammalian cancer cell lines in Table 2.1 were cultured in Dulbecco's Modified Eagle Medium (DMEM) with L-glucose and 0.58g/L L-glutamine (Pan Biotech, #P04-03550), supplemented with 5% foetal bovine serum (FBS) (Pan Biotech, #P30-3031) and with or without a mixture of 100U/mL penicillin and 100ug/mL streptomycin (Pan Biotech, #P06-07100) as required. Cells were maintained at 37°C with 5% CO_2, media change were required every 48-72 hours. The culture was left to grow to required confluency before washed with 1X phosphate-buffered saline (PBS) (ThermoFisher, #BR0014G) and trypsinised using 0.05% Trypsin-EDTA (Pan Biotech, #P10-0235SP). Followed by inactivation of trypsin with DMEM containing FBS. Cells were then spun down at 300 G for 5mins and the resuspended cell pellets were plated into fresh culture flasks.

Cell lines	Cancer/Tissue of origin
CaSki	HPV16+ cervical squamous cell carcinoma
HeLa	HPV18+ endocervical adenocarcinoma
UM-SCC-104	HPV16+ floor of mouth squamous cell carcinoma
93-Vu-147T	HPV16+ floor of mouth squamous cell carcinoma
SiHa	HPV16+ cervical squamous cell carcinoma
PCI30	HPV- tongue squamous cell carcinoma
C33a	HPV- Cervical squamous cell carcinoma

Table 2.1 Mammalian HPV+/HPV- cancer cell lines

2.1.2 NIKS

Mitomycin-C treated Murine 3T3-J2 fibroblast cells (known as feeder cells) were plated out at least 4 hours before they were co-cultured with NIKS cells. The NIKS were maintained in culture media which comprises of Ham's F-12 medium (Pan Biotech, #P04-14500) and DMEM with L-glucose and L-glutamine, at a ratio of 3:1, respectively. To this was added 5% FBS and with or without a mixture of 100U/mL penicillin and 100ug/mL streptomycin as required. The medium was supplemented with 24ug/ml adenine (Sigma, #A2786-5G), 8.3ng/ml cholera toxin (EMD Biosciences #227035), 5ug/ml insulin (Sigma, #I6634-50MG), 0.4ug/ml hydrocortisone (Sigma, #H0888-1G) and 10ng/ml epidermal growth factor (EGF) (Sigma, #E1257-0.1MG).

NIKS were maintained at humidified atmosphere at 37°C with 5% CO_2, media change are necessary every 24-48 hours. To prevent over-confluency, cells were passaged before 80% confluency was reached. The flasks were washed with 1X PBS and incubated in 0.05% Trypsin-EDTA. As NIKS are more adherent to culture flasks than feeder cells, they remain attached to the surface while feeder cells became dislodged and removed. Second round of trypsinisation was performed to disassociate NIKS from tissue culture flasks. The cells were then spun down (at 300 G for 5 minutes) and the pellets were re-plated with newly cultured feeder cells.

2.2 RNA and protein extraction

Cells were harvested by washing the flask once with 1x PBS, followed by incubation with 1 to 3ml of trypsin at 37°C. Required cell media containing FBS were added to the flasks to inactivate trypsin which was collected in 15ml falcon tubes that were spun down at

300 G for 5 minutes. Media were aspirated and the cell pellets obtained were immediately placed on dry ice (known as snap freezing). On occasion where the pellets were required for RNA extraction, tissue culture flasks were directly placed on dry ice after washing and aspirated with cold 1x PBS. Both flasks and cell pellets were stored in -80°C freezer for downstream applications.

Total cellular RNA was isolated from either cell pellets obtained from a T-25 flask or 6-well plates (siRNA experiments) using New England BioLabs Monarch Total RNA Miniprep Kit (NEB, #T2010S), according to the manufacturer's protocol. Cells from 6-well plates were extracted using fresh cell scrapers. RNase-free DNase Kit (Qiagen, #79256) were used for DNA digestion during on-column RNA purification. Eluted RNA samples were stored at -80°C for quantitative polymerase chain reaction (qPCR) analysis.

Protein samples were obtained from T-25 flasks using sterilised cell scrapers and 30-50ul of lysis buffer comprise of M-PER mammalian protein extraction reagent (ThermoFisher, #78501), 25x cOmplete protease inhibitor cocktail (Sigma, #04693116001), 20x PhosSTOP phosphatase inhibitors (Sigma, #04906837001) and Benzonase Nuclease (Sigma, #E1014-5KU) at 0.2unit/ul. Lysis buffer must remain on ice at all time. Cell pellets were directly resuspended in the lysis buffer. Sample were then transferred to a 1.5ml Eppendorf tube and lysed on ice for 20 minutes. Followed by centrifugation at 17,000 G for 15 minutes at 4°C to remove cell debris, and the supernatant containing protein were carefully extracted into a fresh Eppendorf tube.

2.3 RNA analysis

2.3.1 cDNA synthesis

Concentrations of extracted RNA samples were quantified using the NanoDrop One Microvolume UV-Vis Spectrophotometer (ThermoFisher). Initially, up to 5µg/ul of total RNA were converted into complementary DNA (cDNA) using the GoScript Reverse Transcription System (Promega, #A5000), in accordance with the manufacturer's instructions. Later in the project, LunaScript RT SuperMix Kit (NEB, #E3010L) was used instead due to higher conversion efficiency and reduced experiment time. All the experiments included a negative control with no reverse transcriptase present.

2.3.2 Quantitative PCR

All the qPCR reactions were comprised of 5ul PowerUp SYBR Green Master Mix (ThermoFisher, # A25918), 300nM (final concentration) of each forward and reverse primer, 2ul 10-fold diluted cDNA samples and required volume of deionised water to make a total reaction volume of 10ul. The assay was set up in 96-well plates using QuantStudio Design & Analysis software, with the cycle parameters shown in Table 2.2. The genes and primers used for qPCR analysis are shown in Table 2.3.

Stages	Temperature/°C	Time	Cycles
Initial denaturation	95	10 min	1
Denaturation	95	15 sec	40
Annealing	60	1 min	
Melt curve	95	15 sec	1
	60	1 min	
	95	1 sec	

Table 2.2 Cycle parameters for comparative and standard curve qPCR

A negative control (of deionised water instead of cDNA) was included with each experiment. *TATA-binding protein* (*TBP*) was used as the control gene. All reactions were performed in duplicates and were repeated at least three times. Raw data (Ct value) was processed in Excel, where standard curve for each primer set were used to convert raw data into relative expression of transcripts for each gene that was analysed.

Gene	Forward (5'-3')	Reverse (5'-3')	Supplier
TBP	CCCATGACTCCCATGACC	TTTACAACCAAGATTCACTGTGG	IDT
SYCP2	ATACCAGATTCACAGGCAGCGGAA	TGTCCAACATGCCCATTTGCTGTG	IDT
E6	TCAGGACCCACAGGAGCG	CCTCACGTCGCAGTAACTGTT	Eurofin

Table 2.3 Primer sequences for qPCR

2.4 Protein quantification and analysis

2.4.1 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protein concentrations were quantified using the Pierce BCA Protein Assay Kit (ThermoFisher, #23227). Samples were prepared with 4x Laemmli sample buffer (final concentration of 240mM Tris-HCL pH6.8, 8% SDS, 40% glycerol and 0.04% bromophenol blue), 1M dithiothreitol and heated at 90°C for 11 minutes. Followed by SDS-PAGE ran on either 8% or 10% handcast polyacrylamide gels with Tris/Glycine/SDS running buffer containing 25nM Tris, 192nM glycine, 0.1% SDS, (pH 8.3) (BioRad). The gel percentage is an indication for the relative pore size of the polyacrylamide gel. The higher the percentage, the smaller the pores. Electrophoresis was carried out at 150v for 50 minutes and protein samples were transferred from the gel to a Polyvinylidene Difluoride (PVDF) membrane using the Trans-Blot Turbo Transfer System (BioRad) according to the manufacturer's protocol. The run time was extended to 12 minutes for proteins with high molecular wright. Whatman filter paper was shown to facilitate a complete transfer of the bands.

2.4.2 Western blot analysis

The membrane was blocked in 1x Tris-buffered saline (TBS) (pH7.6) with 0.1% Tween-20 and 5% skimmed milk for 1-hr on a rocker at room temperature. Followed by overnight incubation with anti-SYCP2 primary antibody (1:250) (Abcam, #ab138562) on a rocker at 4°C. The PVDF membrane was then washed three times for 10 minutes each with TBS with 0.1% Tween-20 solution. Afterward, it was further incubated with horseradish peroxidase (HRP)-conjugated anti-rabbit secondary antibody (1:10,000) at room temperature for 1-hr with agitation. The membrane was then washed three times

for 10 minutes each with TBS with 0.1% Tween-20. Lastly, it was incubated with enhanced luminol-based detection (ECL) substrate (BioRad, #1705061), followed by protein detection on either a Syngene G:Box-fluorescence imager or using x-ray films.

2.5 Generating standard curve for *SYCP2* quantification

2.5.1 Topo cloning

Two primers targeted at a partial *SYCP2* RNA sequence were designed on Integrated DNA Technology website (Table 2.4) with the aim to cover the transcript region targeted by SYCP2 qPCR primers.

PCR Primers	Forward (5'-3')	Reverse (5'-3')	Product size/bp
Primer set 1	TCAATACCACCACGAAGAAGAA	CAATGTGTCTCCTGACGAAGTA	417
Primer set 2	CGGAGAAGAGCTCGTTAGTTTA	CAGATTCAACAGGCCAATCAC	872

Table 2.4 Primers designed to use in Platinum SuperFi PCR

Both primer sets were used with cDNA samples from the cell line UM-SCC-104 to amplify *SYCP2* gene using Platinum SuperFi PCR Master Mix (Invitrogen, #12358010), in accordance with the manufacturer's instructions. Blunt-end PCR products were generated which were then cloned into a plasmid (pCR-Blunt II-Topo), shown in Figure 2.1 using Zero Blunt TOPO PCR Cloning Kit (Invitrogen, #K2800-20SC), according to the manufacturer's instructions.

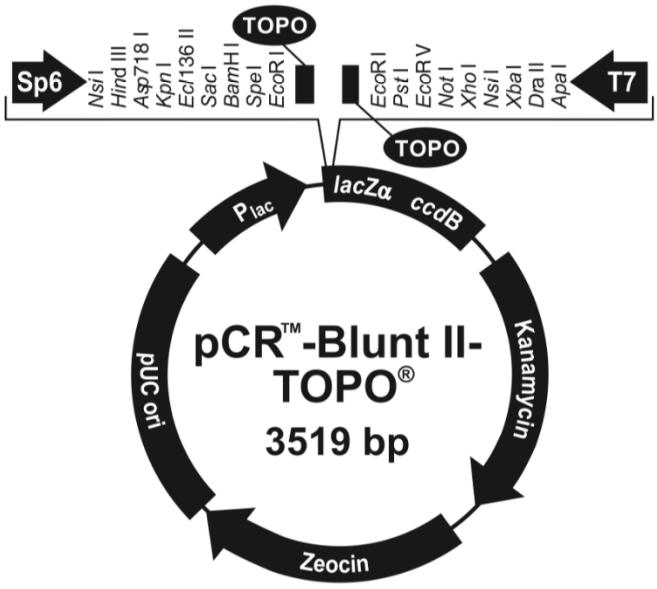


Figure 2.1 Map of pCR-Blunt II-Topo

Genome map of Topo plasmid showing features including *lac* promoter/operator region at bases 95-216, TOPO-Cloning site at bases 336-337, SP6 promoter priming site at bases 239-256, T7 promoter priming site at bases 406-425 and kanamycin resistance ORF at bases 1237-2031 with *kan* promoter positioned at bases 1099-1236.

2.5.2 Bacterial transformation

50ul of *E. coli* DH5-alpha competent cells (NEB, #C2987I) were thawed on ice for 20-30 minutes, followed by addition of Topo plasmid containing *SYCP2* gene. The microcentrifuge tube was incubated on ice for a further 20-30 minutes and then heatshocked for precisely 30 seconds at 42°C. Further 2 minutes on ice was required to enhance plasmid DNA entry. The outgrowth stage was performed with 250ul of SOC media for 1 hour in a 37°C shaking incubator. This enabled *E. coli* cells to establish kanamycin resistance via uptake of the Topo plasmid. Finally, different volumes (50-150ul) of the mixture were plated out onto pre-warmed Luria Bertani (LB) agar plates containing 50µg/ml of kanamycin using sterile technique. The plates were incubated at 37°C overnight. The kanamycin resistance gene in pCR-Blunt II-Topo acted as a selective marker for *E. coli* growth. Only the successfully transformed cells with their expressed resistance gene will result in colonies on the kanamycin agar plates.

2.5.3 Liquid culture and plasmid DNA purification

Following successful transformation, single colonies were picked using sterile pipette tip and inoculated in a falcon tube containing 4ul of LB broth and 4ug/ml of kanamycin. The tubes were placed vertically in a 37°C shaking incubator for 18 hours with the speed of 210rcf. After incubation, bacterial growth was characterised by the opaqueness of the media and plasmid purification was subsequently performed using QIAprep Spin Miniprep Kit (Qiagen, #27106), according to the manufacturer's instructions. Concentration of extracted DNA were quantified using a NanoDrop One Microvolume UV-Vis Spectrophotometer. 15ul of DNA samples at 30ng/ul were sequenced using Sanger sequencing technology using standard primers from IDT, SP6 and T7.

2.5.4 SYCP2 standard curve

Miniprep sample with correct *SYCP2* insert (confirmed by sequencing data) were used to generate *SYCP2* standard curve using qPCR. With given total plasmid size (including *SYCP2* insert) and concentration of the miniprep sample, concentration of plasmid required to have 10^8 copies/ul was determined. Followed by further calculation to result in a master plasmid stock of 10^9 copies/2ul. Seven individual samples were produced using serial dilution (1:10) with plasmid copy numbers going from 10^2 copies/2ul to 10^8 copies/2ul. qPCR cycle parameters used are described in Table 2.2. Experiment was performed in triplicate wells in a 96-well plate and repeated three times. The average threshold cycle (Ct) value was plotted against the dilution factor and the data was fitted to a straight line to give the standard curve. The linear equation of the plot was then used to determine the relative copy level of a given gene in unknown samples (Calculations are shown in 3.1.2).

2.6 RNA interference with siRNA

2.6.1 Reverse transfection of siRNA

Each experiment consists of mock, control siRNAs and *SYCP2* siRNA in triplicates and was repeated at least three times. All siRNAs used were provided by Horizon Dharmacon and are presented in Table 2.5. For each well, 0.2ul of 20uM siRNA stock was diluted in 19.8ul of Opti-MEM (ThermoFisher, #31985062). In mock transfected wells, equal volume of Opti-MEM was added instead of the siRNA. A master mix was made with 0.2ul of Lipofectamine RNAiMAX Transfection Reagent (Invitrogen, #12323563) and 19.8ul of Opti-MEM per well. 5 minutes incubation at room temperature was required to allow the mixing of reagents, followed by the addition of 20ul of Opti-MEM. 40ul of the master mix was added to each well to reach a final siRNA concentration of 20nM. The plate was then shortly spun on PlateFuge Microcentrifuge (Benchmark Scientific), followed by incubation at 37°C for 30 minutes. Cells seeded in a T25 flask with 60-80% confluency were passaged, counted, and added at a concentration of 2000 cells per well in 140ul of DMEM without antibiotics.

On day 5 post transfection (given that day 0 is transfection day), cell viability assay was performed to assess transfection efficiency. Media were aspirated and replaced with 100ul of DMEM/5% FBS without antibiotics. To this was added 20ul of CellTiter 96 AQueous One Solution (Promega, #G3582), followed by incubation at 37°C for 2.5 hours and absorbance reading at 490nm using a CLARIOstar microplate reader (BMG LABTECH). Data normalisation was performed against triplicates of blank wells containing no cells and equal volume of CellTiter 96 AQueous One Solution.

siRNAs	Catalogue number (Horizon)
ON-TARGETplus Non-targeting Control	D-001810-01-05
SMARTpool: ON-TARGETplus SYCP2	L-020632-00-0005
SMARTpool: ON-TARGETplus PLK1	L-003290-00-0005

Table 2.5 siRNAs used in reverse siRNA transfection

2.6.2 Forward transfection of siRNA

Cell suspension from reverse transfection were used for seeding cells for forward transfection. 300,000 (SiHa) – 400,000 (UM-SCC-104 and 93Vu-147T) cells were seeded into a 6-well plate in 2.5ml of DMEM/5% FBS without antibiotics. This occurred 48 hours (SiHa and UM-SCC-104) or 72 hours (93-VU-147T) prior to forward transfection, allowing the culture to reach 60-80% confluency. On the transfection day, 5ul of Lipofectamine RNAiMAX was combined with 245ul of Opti-MEM in a microcentrifuge tube and was allowed to incubate for 5 minutes at room temperature. 3ul of 20uM siRNA stock (negative control and *SYCP2*) was added to 247ul of Opti-MEM in a separate microcentrifuge tube. siRNA was replaced with equal volume of Opti-MEM in mock transfections. They were performed to determine non-specific effects that may have resulted from the transfection reagents or the experiment. The two tubes were combined and incubated at room temperature for 30 minutes. Finally, the mixture is added to the 6-well plates to which a media change was carried out on the day, this results in a final siRNA concentration of 20nM. Cells were harvested 48 hours post transfection by snap freezing the 6-well plates after it was rinsed with cold 1x PBS. RNA extraction was subsequently performed as described in section 2.2 and used in qPCR analysis.

2.7 CRISPR knockout in NIKS

2.7.1 gRNA and vector construct

Five gRNA designs (provided by Horizon) were used in this project (Table 2.6), each targeting different *SYCP2* exon regions as illustrated in Figure 2.2

gRNA catalogue number	gRNA name used	Sequence (5'-3')
SYCP2_147644	gRNA 4	ACCTGGAGATCTGGTCTTAT
SYCP2_147645	gRNA 5	AAATATATTTCTCTAGCAGT
SYCP2_147646	gRNA 6	ATATAAGGTTGTCCACCTTG
SYCP2_147647	gRNA 7	ATTTCAGGAACTTAATAAAG
SYCP2_147648	gRNA 8	CCACAATGTTTCAGCCATTT

Table 2.6 *SYCP2* gRNA sequences

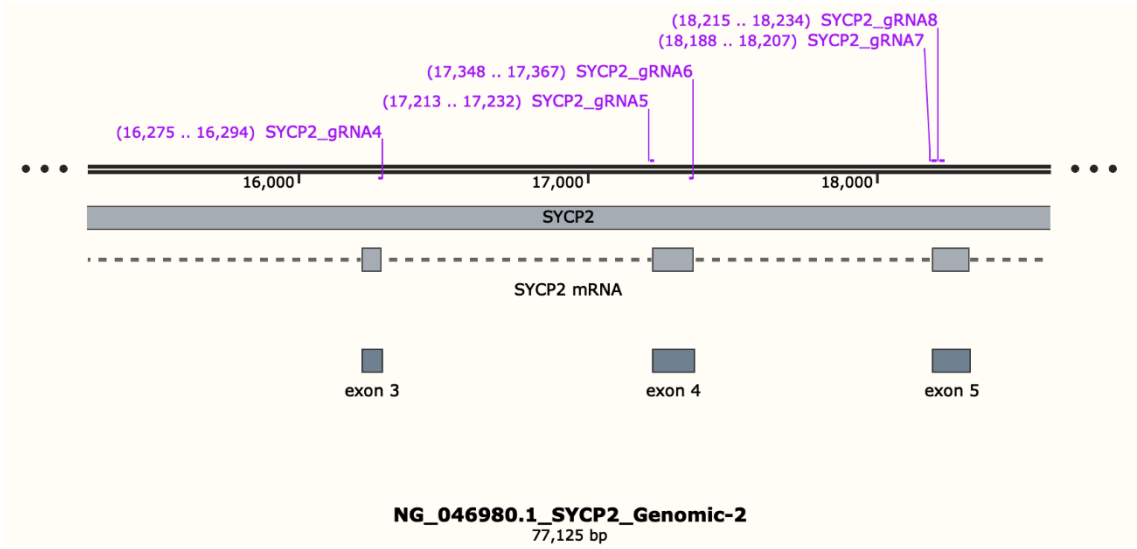


Figure 2.2 Linear map of *SYCP2* genome with gRNA positions

SnapGene revealed that exon 3 is targeted by gRNA 4, gRNA 5 and 6 both target exon 4 and gRNA 7 and 8 target exon 5 of *SYCP2.*

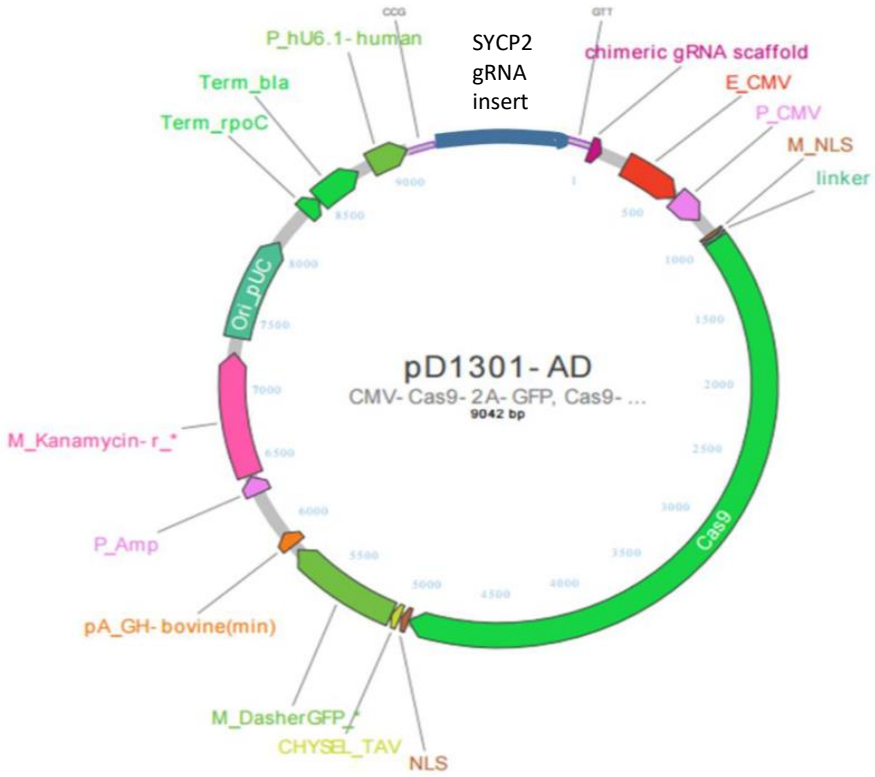


Figure 2.3 Genome map of pD1301-AD with *SYCP2* gRNA insert

The vector provided by Horizon comprises of 9042bp, with CRISPR-Cas9 system and *GFP* gene included in the construct. Ori_pUC acts as the origin of replication. Promoter hU6 regulates gRNA expression and *Cas9* is controlled by promoter CMV. Presence of *GFP* gene allows positive selection for desired insertion and successful transfection. The vector also comprises of an NLS sequence, allowing import of proteins into the cell nucleus. Kanamycin resistance gene allows the vector to be selected for *E. coli* growth.

Vector pD1301-AD (provided by Horizon) was used in this project for the delivery of *SYCP2* CRISPR-Cas9 system (shown in Figure 2.3). It contains two separate promoters, human U6 and cytomegalovirus (CMV), which regulates the expression of *SYCP2* gRNAs, *Cas9* enzyme, *GFP* and kanamycin resistance gene.

2.7.2 NIKS transfection

Wild type (wt) NIKS were maintained as described in section 2.1.2. Typically, the cells were seeded 24-48 hours before transfection, allowing 70-80% confluency on the day of transfection. The experiments were carried out in T25 flasks using FuGENE HD Transfection Reagent (Promega, #E2311) and following the FuGENE HD Technical Manual. Briefly, 18ul of transfection reagent and 6ug of DNA (plasmid containing CRISPR gRNA) was needed per T25 flask with, giving a ratio of 3:1, respectively. Required volume of Opti-MEM was added to provide a total volume of 300ul. The transfection mixture was incubated at room temperature for 15mins before it was added to the T25 flasks containing newly added antibiotics-free FC media. The culture was incubated at 37°C with 5% CO_2 and imaged 48h post transfection using a fluorescence microscope to estimate transfection efficiency.

2.7.3 Flow cytometry

Cell were sorted for GFP signal 48h post transfection using BD FACSJazz flow cytometer (BD Biosciences). Prior to sample preparation, the machine was set up and calibrated, along with correctly aligned lasers and streams. Once the machine is ready, the cells were trypsinised as previously described in section 2.1.2 but the cell pellets were resuspended in 1ml of 1x PBS instead. The samples were loaded into round-bottom polystyrene capped falcon tubes and kept on ice to prevent aggregation. Wt NIKS were used to determine background signal due to cell autofluorescence during sorting. The sample tubes were placed into the FACSJazz machine and the GFP-positive cells were sorted by creating gates outside of the autofluorescence seen in wt NIKS cells. The collected cells were plated into 6-well plates containing a monolayer of feeder cells and

were maintained at 37°C with 5% CO_2. Data obtained were analysed using the BD FACS Sortware software.

2.7.4 DNA extraction

After passaging post-sorted NIKS cells into T25 flasks, they were allowed to grow to 80% confluency before been used to generate frozen stocks and DNA samples. Cell suspensions were divided into two falcon tubes before centrifugation. One tube was resuspended in freezing media containing 10% dimethyl sulfoxide, followed by aliquots into cryovials for storage at -80°C. The other cell pellet was resuspended in 180ul of 1x PBS and DNA extraction was performed using the QIAamp DNA Mini Kit (Qiagen, #51306), in accordance with the manufacturer's instructions. Followed by quantification using a nanodrop spectrophotometer. DNA of wt NIKS was also extracted in the procedure.

2.7.5 Identification of indels with next generation sequencing (NGS)

Requirement for sequencing includes designing primers with incorporated Nextera universal sequencing adaptor, producing amplicons that are approximately 200 base pairs (bp) in length with the CRIPSR cut site located in the middle. PCR amplification of extracted DNA and purification is subsequently performed. Identification of indels were carried out using Illumina Miseq Sequencing at UCL Cancer Genome Engineering Facility and further sequencing analysis were performed using CRISPR-PARSR pipeline created by Dr Javier Herrero from UCL Cancer Institute. Finally, a gene editing report is produced containing information on percentages and locations of indels and the genome editing efficiency of individual gRNAs.

2.7.5.1 Primers design

Suitable forward and reverse primers with an overhang universal adaptor sequence at its 5'-end (illustrated in Figure 2.4) were designed using SnapGene software. All primers were tested for formation of dimers and hairpin/secondary structures using an online primer structure analysis tool on the IDT website (http://eu.idtdna.com/calc/analyzer). Sequences of primers are shown in Table 2.7.

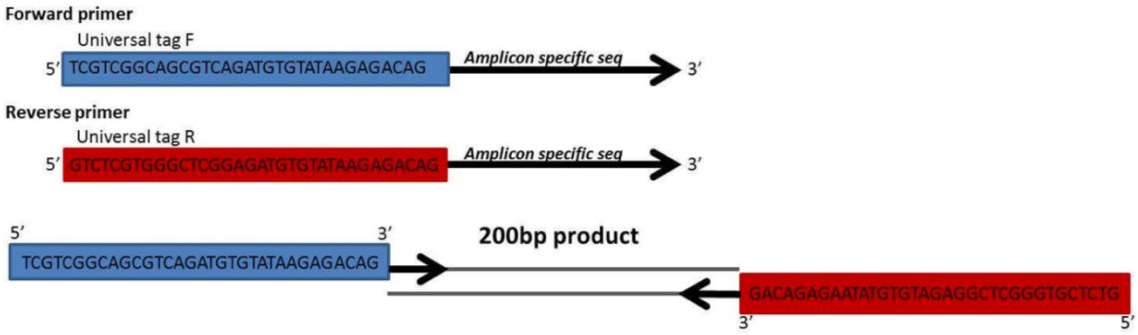


Figure 2.4 Sequence of overhang universal adaptors

Primers	Forward (5'-3')	Reverse (5'-3')	Amplicon size/bp
Primer 1 (for gRNA 4)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAAACAGATTGTGAGATACAG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACATAGCACTGAATACC	195
Primer 2 (for gRNA 5)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGTTCATTTGAGAGTTTTC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GCTGCATTTAATCTTCACATC	214
Primer 3 (for gRNA 6)	TCGTCGGCAGCGTCAGATGTGATAAGAGACAGGCATTGATGATGCTTTAAG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTAATGGAGTACTTCCTTC	208
Primer 4 (for gRNA 7 and 8)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGTGTAACACTCCTTGTAC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GTTTTATCATCGTTAGAAGTCC	192

Table 2.7 Primer designs with universal adaptor and corresponding amplicon sizes

2.7.5.2 PCR amplification

Using the primers in Table 2.7 and extracted DNA samples, amplicons were produced using KAPA HiFi HotStart ReadyMixPCR Kit (KAPABiosystems, #KK2600). The volume of components was scaled accordingly to give a final volume of 10ul, as presented in Table 2.8. PCR conditions used in this project is described in Table 2.9. A gradient of annealing temperature was used to determine the optimised condition for each amplicon. Total of four primer set were used with five DNA samples, generating four amplicons with sizes presented in Table 2.7.

Components	Volume/ul	Final concentration
2X KAPA HiFi HotStart ReadyMix	5	1X
10uM forward primer	0.3	0.3 uM
10um reverse primer	0.3	0.3 uM
DNA	As required	100ng/ul
PCR-grade water	As required	N/A

Table 2.8 KAPA PCR mixture

Stages	Temperature/°C	Time	Cycles
Initial denaturation	95	3 min	1
Denaturation	98	20 sec	30
Annealing	58-62	15 sec	
Extension	72	15 sec	
Final extension	72	1 min	1

Table 2.9 PCR cycle parameters

2.7.5.3 Generation and purification of amplicons

Following PCR reactions, gel electrophoresis was performed with 1ul of product to establish initial assessment of amplification specificity. 1ul of 6x DNA loading dye was added to the product and loaded onto 2% DNA agarose gel with 100bp plus DNA ladder (NEB). Gel electrophoresis was performed at 100V for 32-35 minutes. Upon detection of single bands with corresponding amplicon sizes, the remaining 9ul of PCR product was combined with 2ul of 6x loading dye and gel electrophoresis was repeated. Each band were then cut out and purified using Monarch DNA Gel Extraction Kit (NEB, # T1020L), according to the manufacturer's instruction. 1ul of purified sample were ran on the gel to show single product with no other bands due to primer dimer formation, a gel image was attached with the NGS request form. Purified DNA samples were quantified using a nanodrop spectrophotometer and were submitted for NGS with a total volume of 20ul at 20ng/ul.

3. Results

3.1 Establishment and optimisation of quantification assay for *SYCP2*

3.1.1 Successful gene cloning

Miniprep samples from eight liquid cultures were amplified via PCR using SYCP2 qPCR primers (Table 2.3), followed by gel electrophoresis. This allowed the initial assessment of gene cloning success. A single band was detected around 100bp for all miniprep samples, corresponding to the qPCR product size of 105bp for *SYCP2*. Subsequently, the samples were sequenced, and the results obtained indicated successful gene cloning for three out of the eight samples (Figure 3.2). Sequencing data for M1 (primer set 1) is shown in Figure 3.2b where there was an exact sequence match with the product that is generated by primer set 1 (Figure 3.2a). Hence, M1 (primer set 1) was used to carry out standard curve generation via qPCR assay.

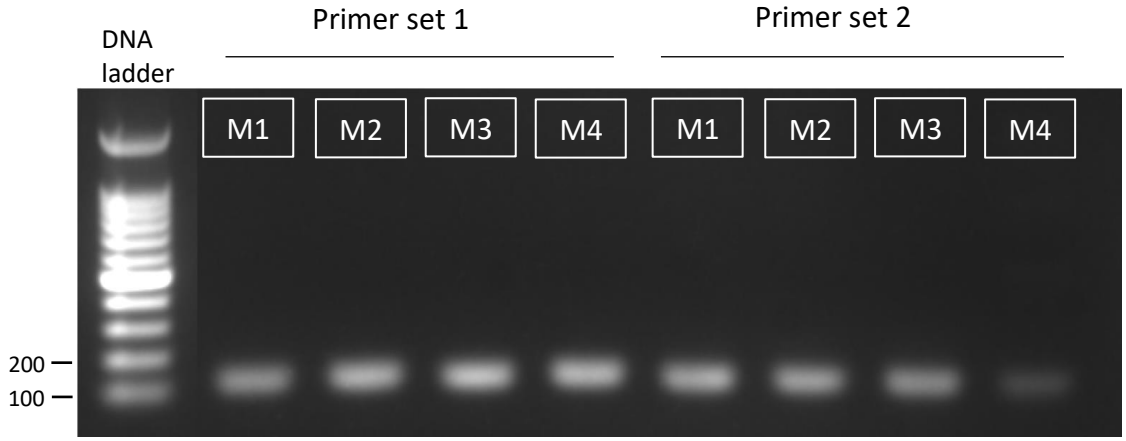


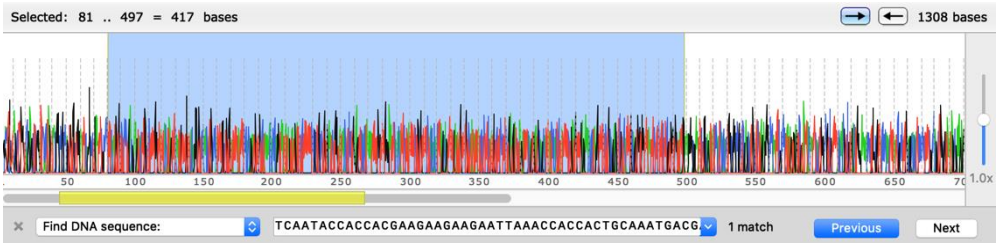
Figure 3.1 Gel electrophoresis of miniprep samples resulted from Topo PCR cloning

Four miniprep samples (denoted as M) were obtained for each primer set designed (Table 2.4). Gel electrophoresis was performed (2% gel, 100V for 35mins) after PCR amplification using 1ul of PCR product and 1ul of 6X DNA loading dye. 5ul of 100bp DNA ladder was loaded in lane 1. Single bands were detected around 100bp for all samples.

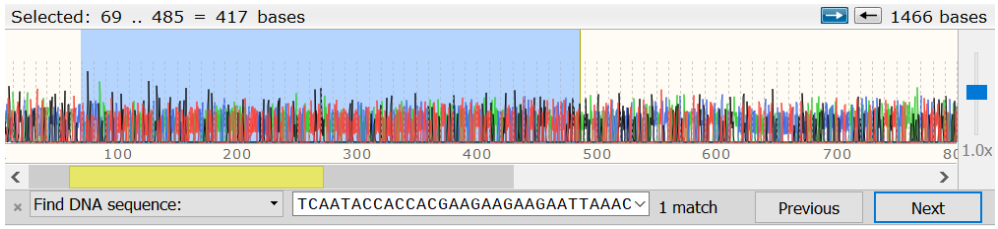
(a)

TCAATACCACCACGAAGAAGAAGAATTAAACCACCACTGCAAATGACGAGCTCTGCAGA
GAAACCTAGTGTTTCTCAAACATCAGAAAATAGAGTGGATAATGCTGCATCACTGAAATCT
AGATCATCAGAAGGAAGACATAGAAGAGATAATATAGACAAACATATCAAAACTGCTAA
GTGTGTAGAAAACACAGAAAATAAGAATGTTGAATTCCCAAACCAAAATTTTAGTGAACT
CCAGGATGTTATACCAGATTCACAGGCAGCGGAAAAAAGAGATCATACTATATTACCTGG
TGTTTTAGACAACATCTGTGGAAATAAAATACACAGCAAATGGGCATGTTGGACACCTGT
AACAAACATTGAACTATGTAATAACCAAAGAGCAAGTACTTCGTCAGGAGACACATTG

(b)



(c)



(d)

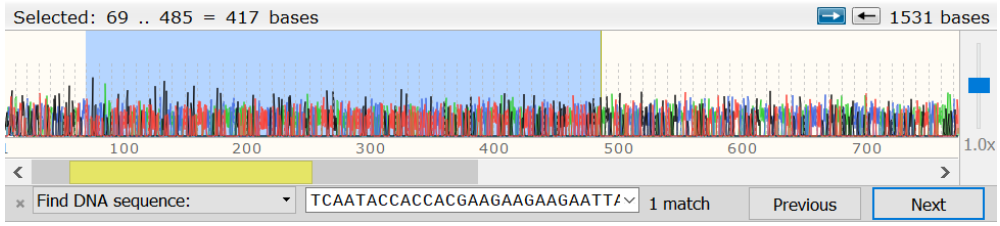


Figure 3.2 Illustration of partial *SYCP2* sequence and corresponding sequencing data

Partial *SYCP2* sequence is shown with (a) red and blue nucleotides corresponding to primer set 1 in Table 2.4, giving a PCR product size of 417bp. The highlighted green and purple sequences refer to the forward and reverse SYCP2 qPCR primers, respectively (Table 2.3). In (b), sequencing data of M1 (primer set 1) using SP6 primer revealed exact match of nucleotides seen in (a), indicating successful gene cloning. Data in (c) and (d) refers to M3 and M4 (primer set 1) from Figure 3.1, respectively.

3.1.2 SYCP2 standard curve

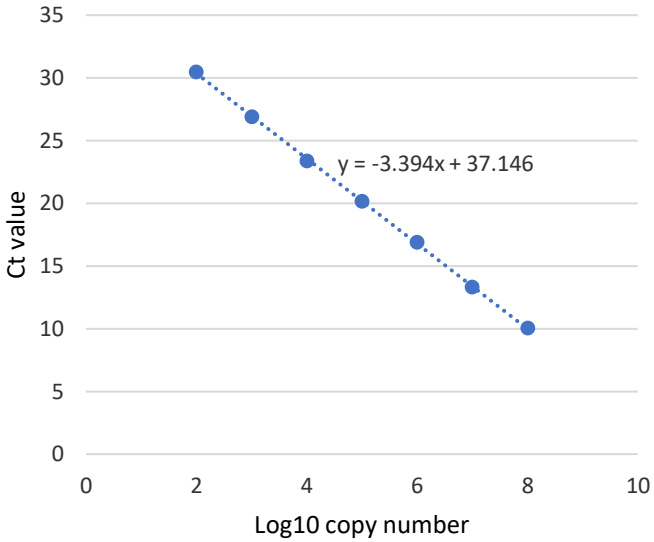


Figure 3.3 Standard curve for SYCP2 qPCR primers

Linear equation (y=-3.394x + 37.146) was generated using average Ct values from repeated qPCR experiments (n=3) with error bars representing SEM.

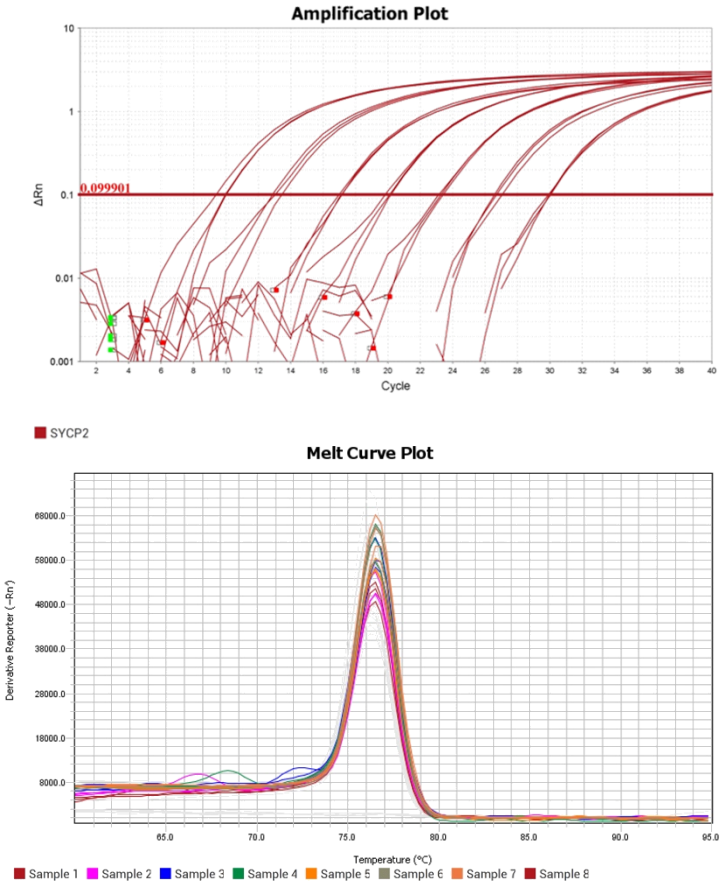


Figure 3.4 Amplification and melt curve plot for miniprep sample using qPCR

The amplification plot revealed a calculated threshold level of 0.099901, corresponding to the cycle number in which a significant amount of *SYCP2* transcript was accumulated. Sample 1 to 7 in the melt curve plot corresponds to the different plasmid concentrations, ranging from 10^2 copies/2ul to 10^8 copies/2ul, respectively. Sample 8 was included as a negative control.

Triplicate repeat of standard curve qPCR assay was performed with M1 (primer set 1) to generate a standard curve for *SYCP2* quantification at the mRNA level. A repeated qPCR assay data is shown in Figure 3.4 which demonstrated the data points used for standard curve was reliable as the technical repeats for each plasmid concentration gave identical or similar Ct values. The melt curve plot illustrated correct amplification of *SYCP2* transcript in samples with different plasmid concentration. High primer efficiency was demonstrated for all three targeted genes analysed in this project (Table 3.1).

The linear equation from Figure 3.3 was used to convert raw data (Ct value) into *SYCP2* transcript copy number. The following calculation was performed using Excel:

⇒ y(Ct) = a + b * x (log10 copy number) a = y-intercept, b = slope gradient

⇒ Log10 copy number = (Ct – a)/b

⇒ Copy number = 10^ (log10 copy number)

E6 standard curve was obtained by carrying out qPCR assay as described in section 2.5.4, using plasmid (pLXSN, 5874bp) with *E6* insert (477bp) provided by Dr Tim Fenton. Standard curve for TBP was generated by Dr Nerissa Kirkwood, member of Tim Fenton's research group. Linear equations for SYCP2, TBP and E6 primers are presented in Table 3.1 and were used throughout this project to calculate SYCP2 and E6 expression level relative to TBP.

qPCR primer targeted genes	Linear equation	Primer efficiency/%
TBP	y = -3.217x + 36.555	104
SYCP2	y = -3.394x + 37.146	99
E6	y = -3.611x + 38.404	89

Table 3.1 Linear equation of standard curve for targeted genes and primer efficiency

3.2 Validating SYCP2 overexpression

SYCP2 overexpression was established via comparative qPCR, using housekeeping gene *TBP* expression as a control. A threshold level was calculated by the QuantStudio Design & Analysis software and the results obtained were given in Ct (threshold cycle) values. This reflects the cycle number at which a significant number of amplicons has accumulated, providing sufficient fluorescence to cross the threshold value. Average Ct values from triplicate repeats were converted into expression values using the standard curve generated for SYCP2 qPCR primers and are shown in relative to TBP expression level.

HPV16 positive cell lines all demonstrated increased SYCP2 expression levels relative to TBP (Figure 3.5a), with 93-Vu-147T cell line exhibiting the highest value of 2.7 (copy number relative to TBP). Notably, SYCP2 expression in HeLa cells (HPV18 positive) was not higher than that seen in the HPV negative cell lines, PCI-30 and C33a, suggesting SYCP2 upregulation is only seen with certain high-risk HPV types. SYCP2 expression levels observed in HPV+ cell lines may be low compared to E6 copy number relative to TBP (Figure 3.5b) but nevertheless, approximately 16- to 30-fold increase was demonstrated when compared to the HPV- cell lines, with 93-VU-147T showing almost 30-fold increase compared to PCI-30. One way ANOVA test showed a statistically significant difference for relative SYCP2 expression level between all HPV16+ cell lines and PCI-30. All of which resulted in a p value of <0.0002.

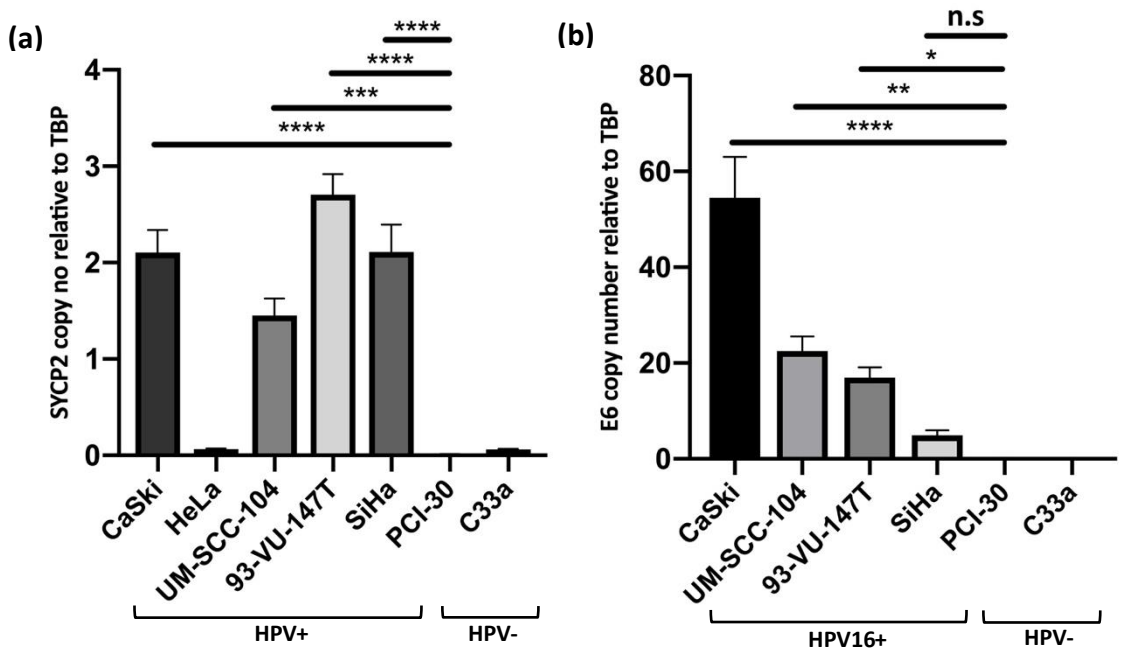


Figure 3.5 Expression of SYCP2 and E6 relative to TBP in RNA samples extracted from HPV+/HPV- cancer cell lines

Average expression of (a) SYCP2 and (b) E6 relative to TBP were gathered from triplicate repeats of qPCR. One way ANOVA statistic tests were performed between HPV+ cell lines and PCI-30 in both set of graphs. *Columns,* mean; *bars*, standard error of mean (SEM).

Further qPCR assays were carried out in HPV16+ and HPV- cell lines with E6 as the gene of interest (Figure 3.5b), as an approach to observe if the expression level of SYCP2 was related to E6 expression. It is clear that E6 expression level is significantly higher than SYCP2 across all HPV16+ cell lines, with CaSki, UM-SCC-104, 93-VU-147T and SiHa showing over 550-, 220-, 160- and 45-fold increase compared to PCI-30, respectively. One way ANOVA test demonstrated significant difference for CaSki ($p < 0.0001$), UM-SCC-104 ($p < 0.01$) and 93-VU-147T ($p < 0.04$) but not SiHa. As expected, the HPV- cells lines do not show any E6 expression. From these results, we can see that the level of SYCP2 expression is not directly correlated with the level of E6 expression HPV16+ cancer cell lines; rather, it would appear that all of the HPV16+ lines, regardless of their E6 expression level, have strongly upregulated SYCP2 expression.

3.3 Effect of siRNA gene silencing on cell viability

3.3.1 MTS assay for siRNA transfection efficiency

The established ectopic expression in HPV+ cancers has led to an assumption of possible dependency between cell survival and SYCP2 levels. Three cell lines (93-VU-147T, UM-SCC-104 and SiHa) with relatively high SYCP2 expression (see Figure 3.5a) were selected to undergo siRNA transfection. To assess the cellular effect of gene silencing, cell viability assay (MTS) was performed on day 5 of reverse siRNA transfection (given transfection occurs on day 0) in 96-well plates. It is a colorimetric assay that is based on the ability of metabolic active cells to reduce the tetrazolium dye 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) to a purple coloured formazan, thereby reflecting the number of viable cells present. The absorbance was read at 490nm using a spectrophotometer, providing quantitative data.

Inhibits of polo-like kinase 1 (PLK1) protein have detrimental effects on cell cycle progression, leading to apoptosis and disruption of mitosis in cancers. Hence *PLK1* siRNA was used as a positive control in this experiment. Non-targeting (NT) siRNA designed not to target any known mRNAs, was used as negative control to allow detection of non-specific effects.

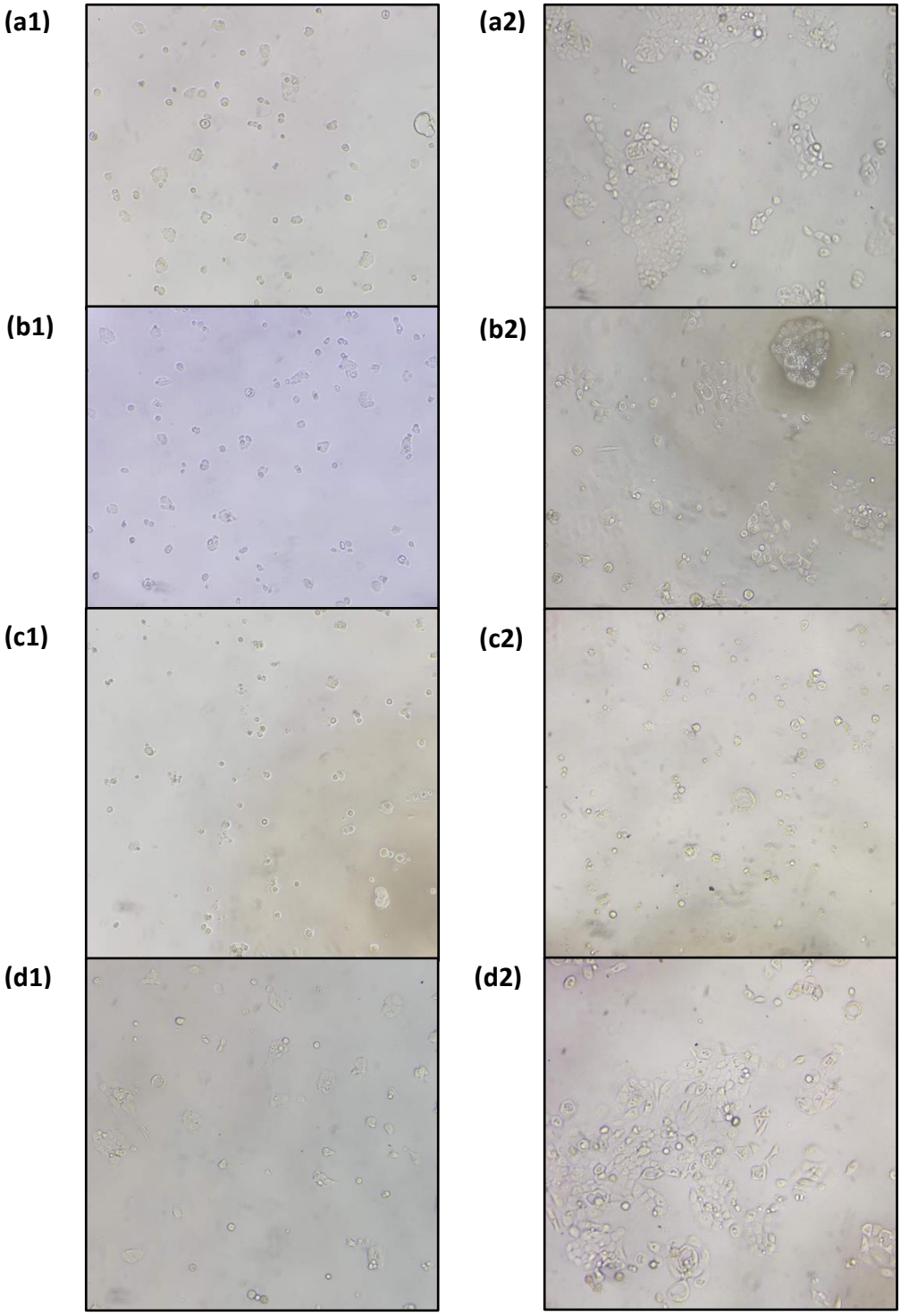


Figure 3.6 Microscopic images of siRNA transfected UM-SCC-104 cells on day 2 and day 5 post transfection

UM-SCC-104 cells were imaged at 100x magnification on day 2 and day 5 post transfection. Figure (a) to (d) represent mock, NT, *PLK1* and *SYCP2* siRNA-transfected cells, respectively, with day 2 designated as (1) and day 5 as (2). Significant low number of viable cells can be seen in the *PLK1*-transfected population compared to other siRNA transfected cultures. Viability does not seem to be affected in cells transfected with *SYCP2* siRNA.

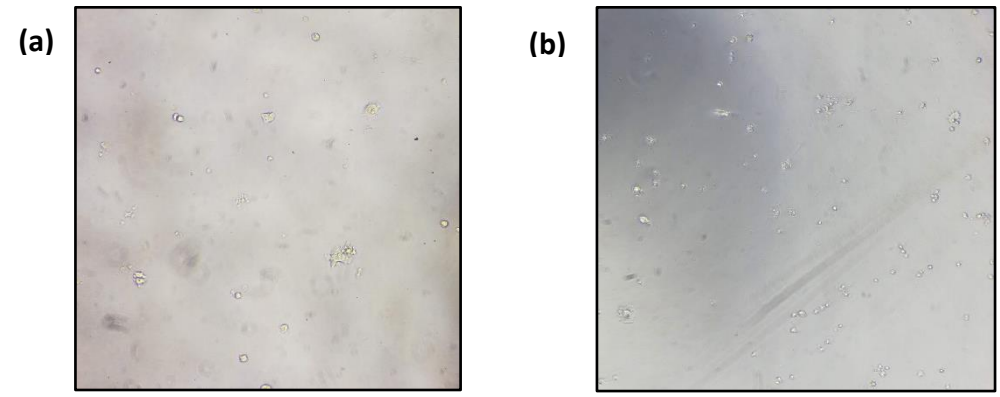


Figure 3.7 Microscopic images of PLK1 transfected 93-VU-147T and SiHa cells on day 5 post transfection

On day 5 post transfection, (a) 93-VU-147T and (b) SiHa cell lines showed significant low number of viable cells in the *PLK1*-transfected population.

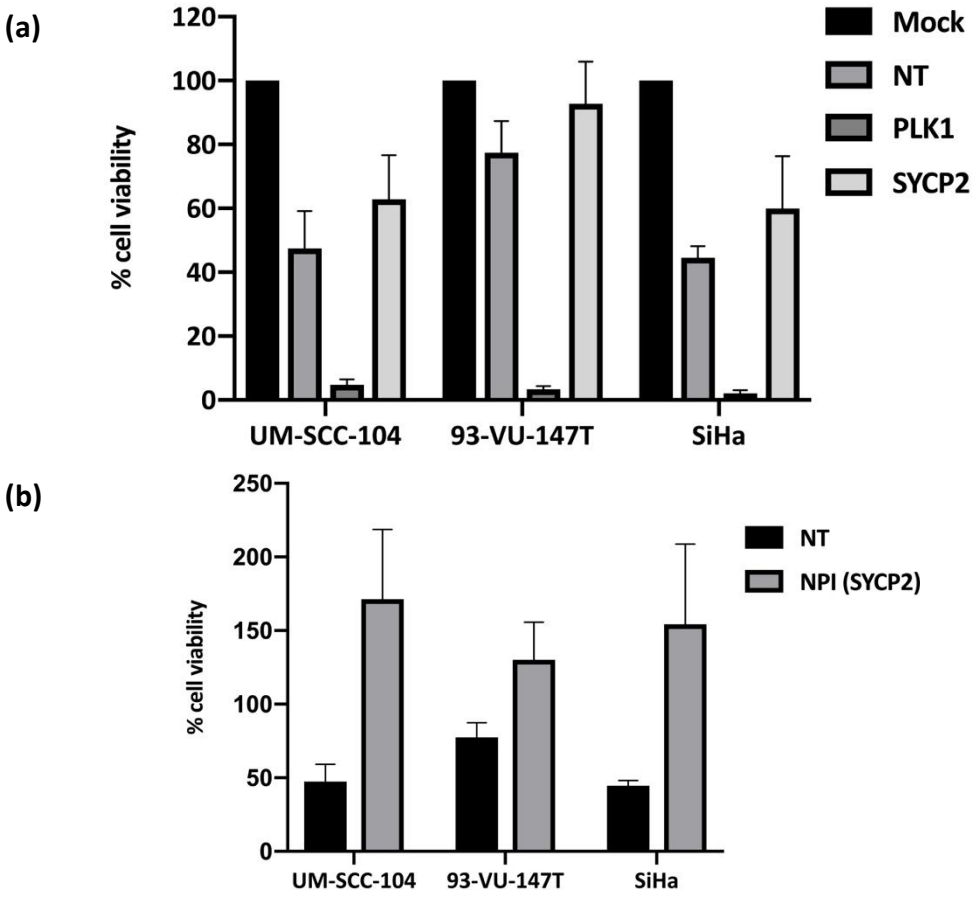


Figure 3.8 Percentage of viable cells post-siRNA transfection using MTS assay

Illustration of viable cell percentages using MTS assay on day 5 post transfection. (a) Viability of three HPV16+ cell lines transfected with mock, control and *SYCP2* siRNAs are shown. In (b), SYCP2 data were normalised with normalised percent inhibition (NPI) calculation, using NT as the upper bound and PLK1 as the lower bound value. *Columns*, mean; *bars*, SEM.

Microscopic images of 93-VU-147T and SiHa (Figure 3.7) cell lines also demonstrated a low number of viable cells in the *PLK1*-silenced populations, as shown for UM-SCC-104 cell line in Figure 3.6. In addition, average of <5% cell viability was calculated for *PLK1*-silenced cells for all three HPV+ cell lines (Figure 3.8a). This result suggested that siRNA transfection efficiency was reasonably high in the experiments performed. Viability for *SYCP2*-silenced cells remained high, surpassing the value for cells transfected with NT siRNA. SYCP2 data was normalised against NT and PLK1 as upper and lower bound, respectively (Figure 3.8b). Resulting NPI data also demonstrated high viability percentages for *SYCP2*-transfected cells compared to the population transfected with NT siRNA.

3.3.2 Confirmation of gene silencing via qPCR

RNA samples with insufficient concentration was extracted from reverse siRNA transfected cells in 96-well plates, leading to the application of 6-well plates using forward siRNA transfection (as described in 2.6.2). In this experiment, components were scaled up to give a final siRNA concentration of 20nM. The results (Figure 3.9) showed that approximately 50% gene silencing was achieved in all three cell lines, by comparing the relative SYCP2 expression level between mock and *SYCP2*-silenced cells. One-way ANOVA statistic test indicated that 93-VU-147T showed the most significant knockdown of *SYCP2* gene. Some significant difference was detected in SiHa, whereas UM-SCC-104 showed no significant difference between mock and *SYCP2*-silenced cells.

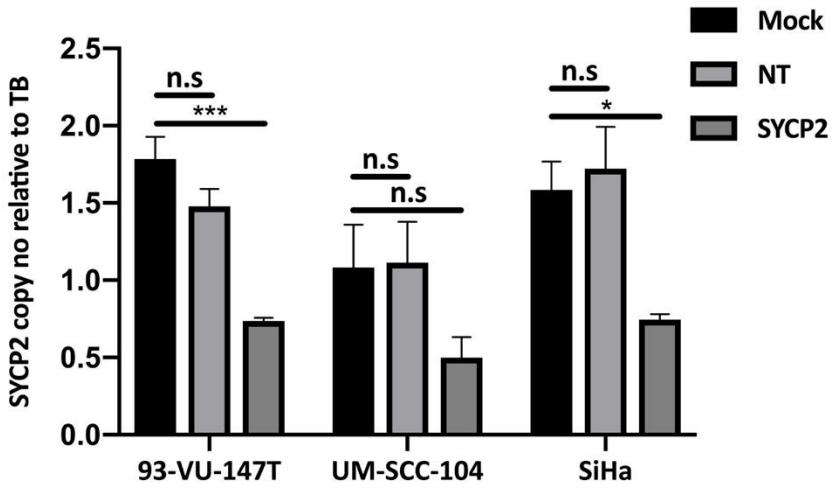


Figure 3.9 Relative SYCP2 expression in three siRNA transfected HPV+ cell lines

RNA samples were extracted from transfected cells in 6-well plates 48h-post transfection, followed by cDNA synthesis and qPCR. Data are shown from repeated (n=4) forward siRNA transfection. One-way ANOVA statistic test was performed between mock and NT/SYCP2 values, no significant difference was observed between the mock transfected and NT transfected cells in all three cell lines. 93-VU-147T showed the most significant difference between mock and *SYCP2* transfected cells with a value of $p < 0.001$. UM-SCC-104 showed no statistic significant ($p < 0.1$) and SiHa had a value of $p < 0.04$, indicating some significant difference. *Columns*, mean; *bars*, SEM.

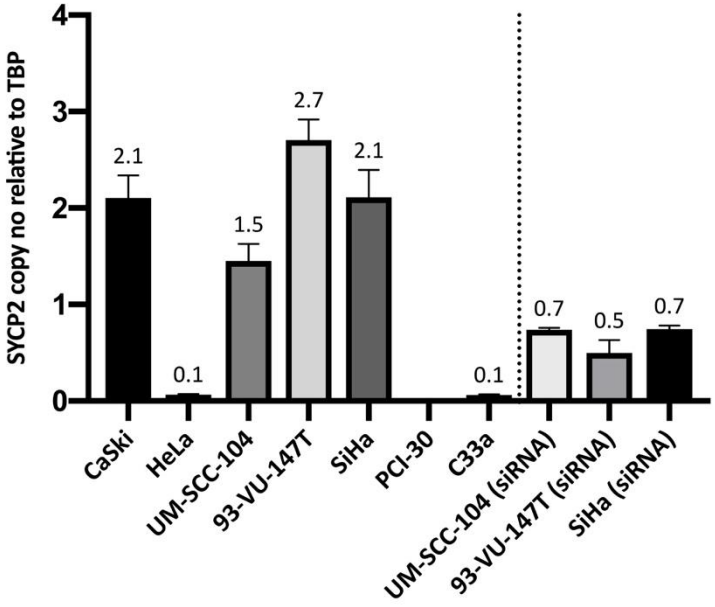


Figure 3.10 Relative SYCP2 expression level in HPV+/HPV- cell lines and siRNA transfected HPV16+ cell lines

Comparison of SYCP2 expression level (relative to TBP) between HPV+/HPV- cell lines and the HPV16+ cells lines transfected with *SYCP2* siRNA. The average values are shown above each bar.

The effect of gene silencing using siRNA was undeniably noticeable as presented in Figure 3.9, where the relative SYCP2 expression seen in *SYCP2*-silenced cells substantially decreased. However, full *SYCP2* knockdown was not achieved in this experiment as the relative expression level was still above (with 5-7 fold differences) HPV- cell lines with no SYCP2 expression (Figure 3.10).

3.4 Generation of Indels using CRISPR-Cas9 system

3.4.1 Purity and sequences of gRNAs

Concentration for each gRNA-containing plasmid was determined using a nanodrop spectrophotometer. Measurements of A260/A280 ratio above 1.80 and A260/A230 above 2.0 indicated DNA purity for all samples (Table 3.2).

SYCP2 gRNA designs	Concentrations(ng/ul)	A260/A280	A260/A230
gRNA 4	85	1.86	2.23
gRNA 5	95	1.87	2.22
gRNA 6	95	1.86	2.28
gRNA 7	91	1.86	2.26
gRNA 8	83	1.86	2.27

Table 3.2 DNA concentration and purity of Horizon SYCP2 gRNAs miniprep

Confirmation of gRNA sequences was achieved through Sanger sequencing (Figure 3.11). The data showed perfect match against the reference sequences in Table 2.6, therefore, CRISPR knockout experiment was proceeded using these five gRNAs.

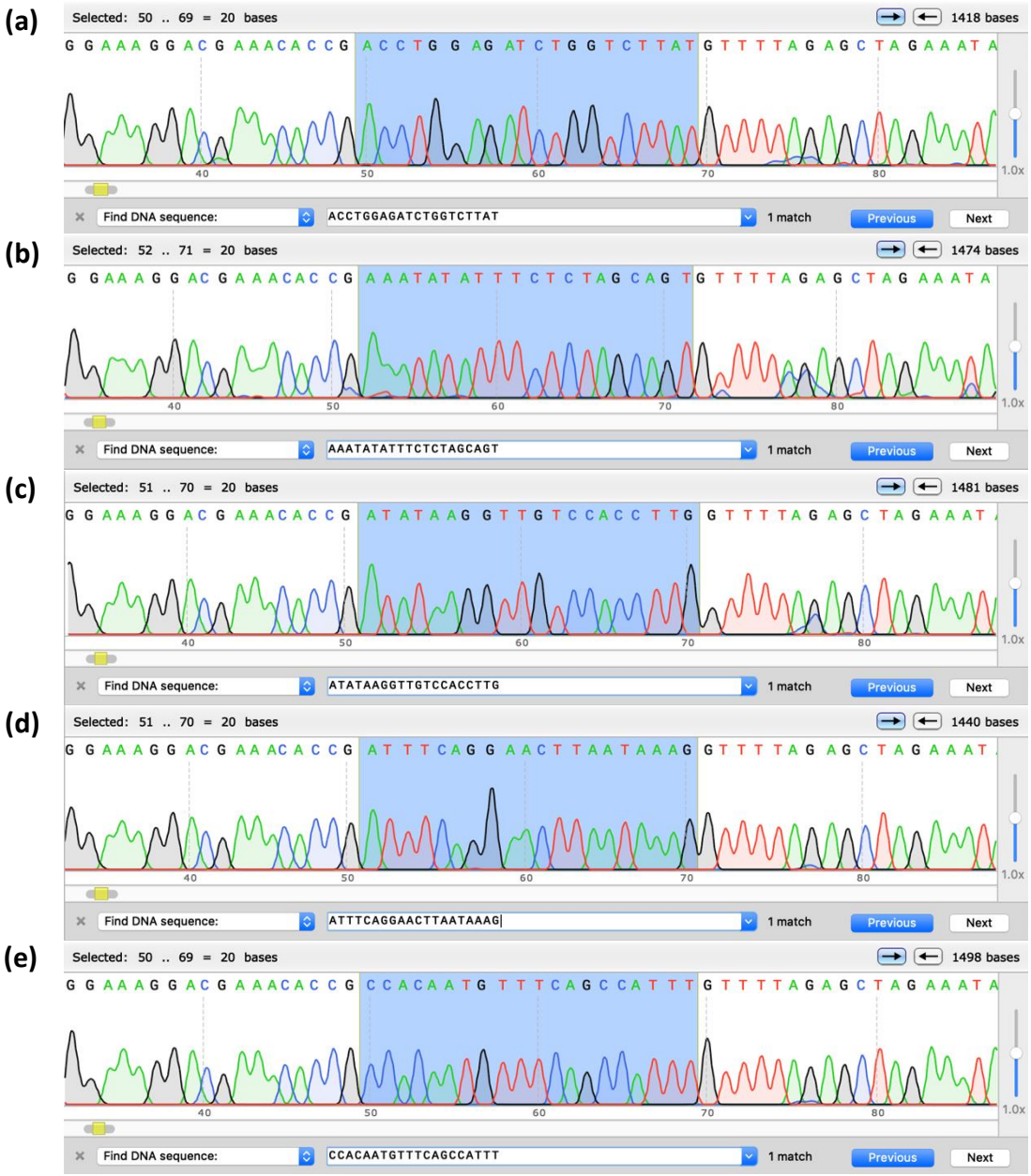


Figure 3.11 Chromatogram data of gRNA sequences

The data confirmed the correct sequence for all five gRNA designs, with (a) to (e) representing *SYCP2* gRNA 4 to 8, respectively.

3.4.2 FACS

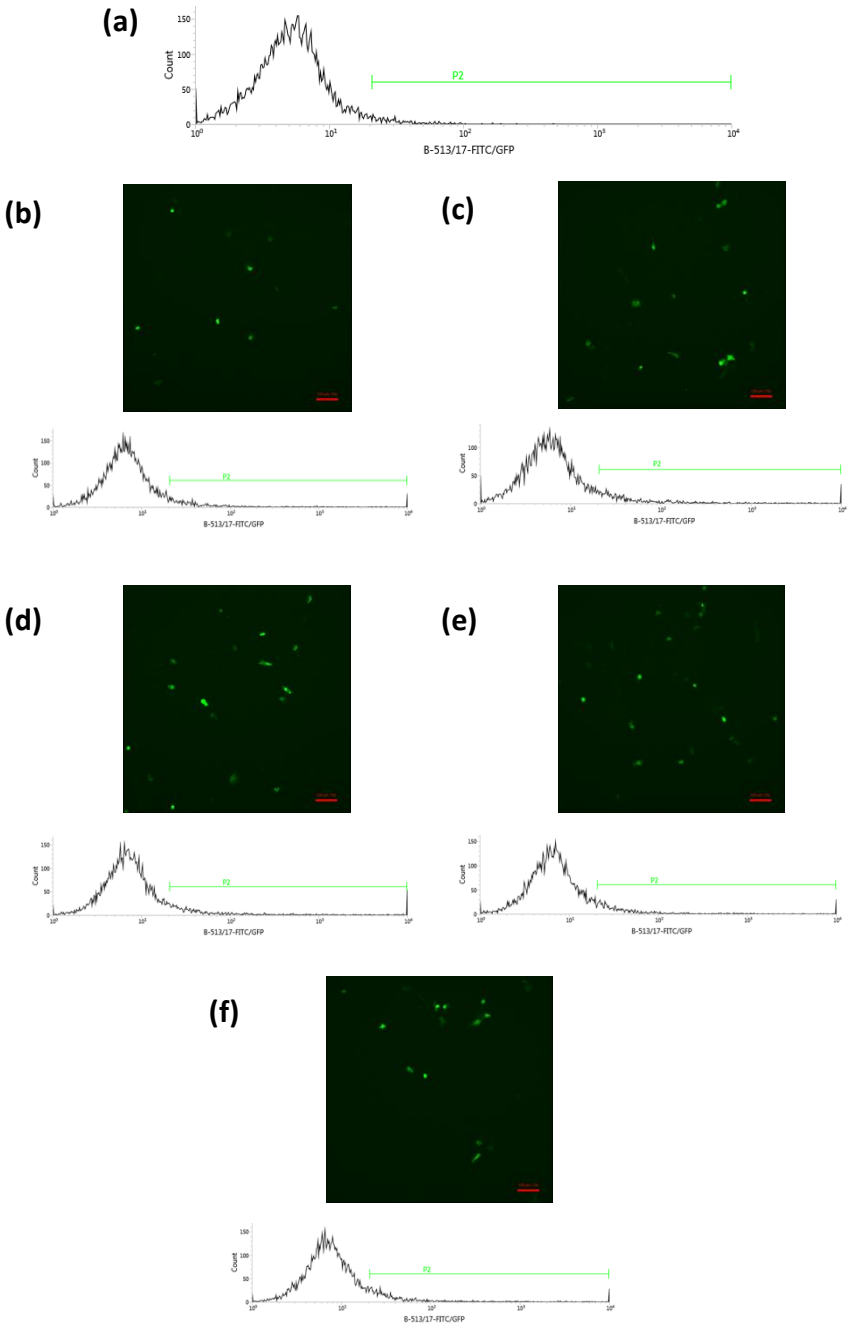


Figure 3.12 Evaluation of CRISPR transfection in NIKS under fluorescence microscope and assigned gates in FACS

Wt NIKS was used in flow cytometry (a) to establish background signal due to autofluorescence and figure (b) to (f) represents gRNA 4 – 8 transfected cell population, respectively. The upper panel shows fluorescence microscope image taken 48hr post transfection and the bottom panel shows the gate that was assigned for sorting GFP-positive cells. Approximately 10,000 events occurred with each cell sample. Green cells were not detected in wt NIKS cells (not shown) but are visible in gRNA transfected cell populations. Cells with fluorescence signal outside of the autofluorescence seen in wt NIKS (indicated by P2) were sorted and collected. A noticeable peak was detected towards the end of the graph (lower panel) for all transfected cells but not in wt NIKS.

The fluorescence microscope images identified many GFP-positive cells under blue light for all gRNA transfected cell populations (Figure 3.12), suggesting great success in delivery of CRISPR-Cas9 system in NIKS cells. However, a peak towards the end of the graph was visible in all transfected populations (lower panels of Figure 3.12) but was not detected in the wt population. Possibly due to the voltage been set too high during sorting, which caused the GFP-positive cells to pile up at the far end of the scale.

Cells with GFP-positive characteristics were sorted/collected and plated into 6-well plates, allowing recovery, growth, and DNA extraction of edited cell population. However, it is not guaranteed that the cells collected were all effectively inserted with the CRISPR construct. Further sequencing analysis will be required to confirm successful transfection.

3.4.3 Amplification and purification of DNA from edited cell population

PCR were performed with parameters from Table 2.9 to establish the optimised annealing temperature for each amplicon product. It was concluded that 60°C was best for amplicon 1 and 58°C for the other three set of amplicons generated from gRNA transfected cells, as determined in Figure 3.13. All four amplicons were also produced from wt NIKS cells (with annealing temperature of 58°C), showing single band around 300bp (Figure 3.13 and 3.14).

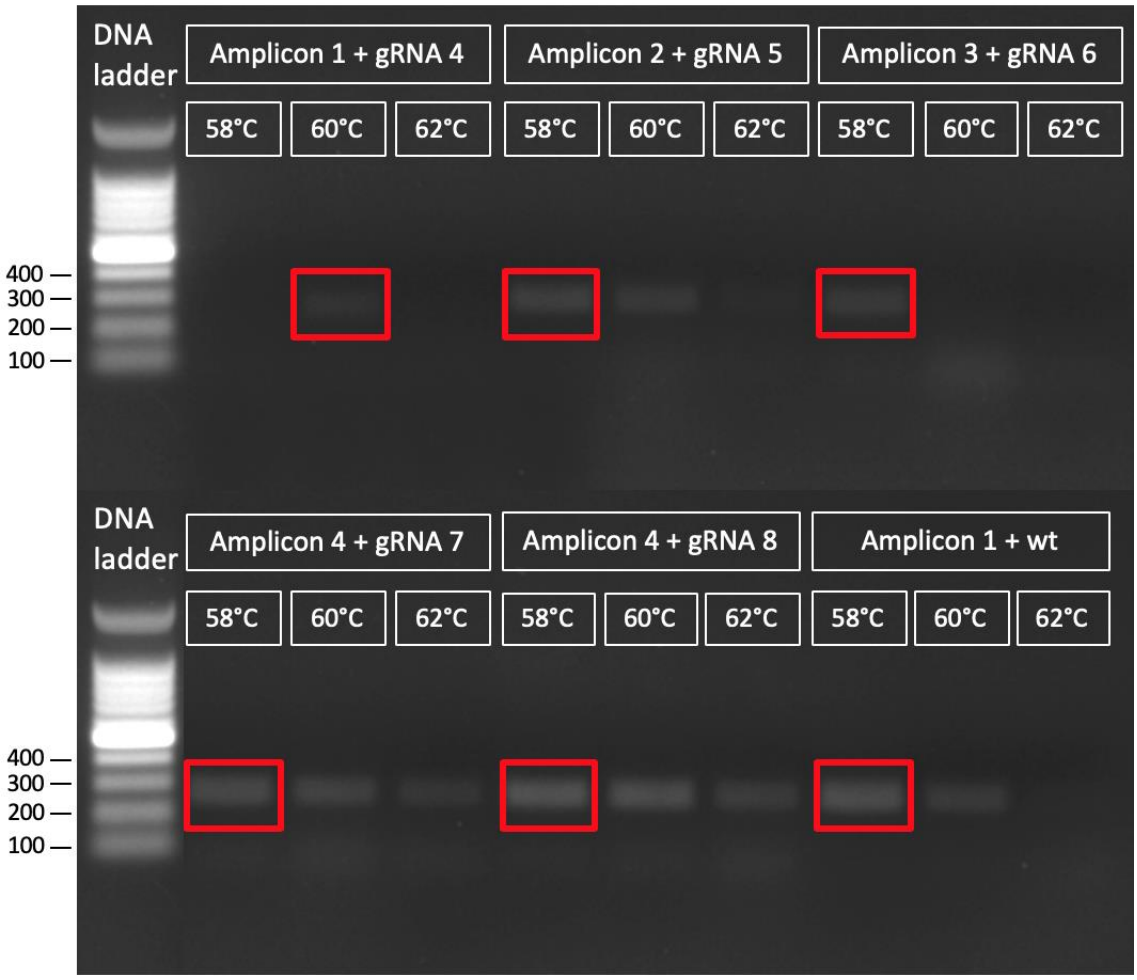


Figure 3.13 Gel electrophoresis of amplicon products from DNA amplification

Total of four set of amplicons were produced with gRNA 7 and 8 producing the same amplicon. DNA from wt NIKS were amplified with primers for amplicon 1. 1ul of PCR product with 1ul of 6x DNA loading dye were loaded into each well on a 2% agarose gel. It ran for 35mins at 100V with 5ul of 100bp DNA ladder in lane 1. Single band between 200 and 300bp was detected in all amplicons. The brightest band (red box) corresponds to the optimised annealing temperature for each amplicon.

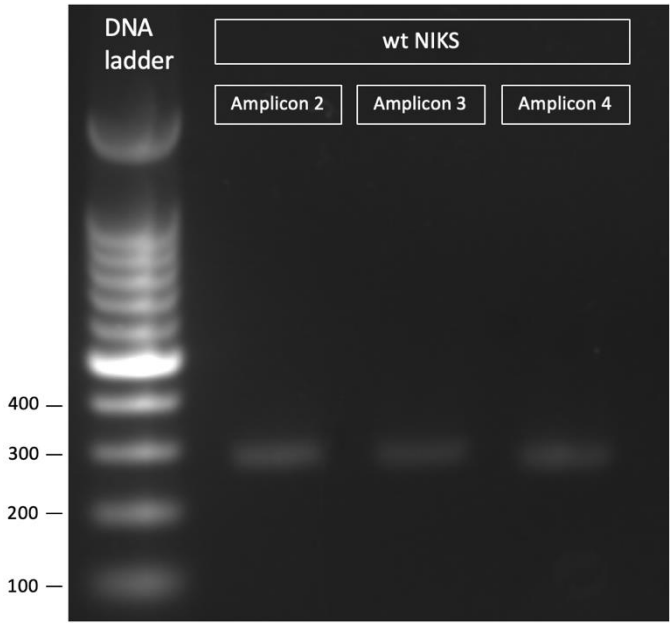


Figure 3.14 Gel electrophoresis of amplicon products from wt NIKS

Amplicons 2 – 4 were amplified using wt NIKS DNA samples with an annealing temperature of 58°C. 1ul of PCR product with 1ul of 6x DNA loading dye were loaded into each well on a 2% agarose gel with a running time of 35mins at 100V. 5ul of 100bp DNA ladder was loaded in lane 1. Single band was detected for all amplicons around 300bp.

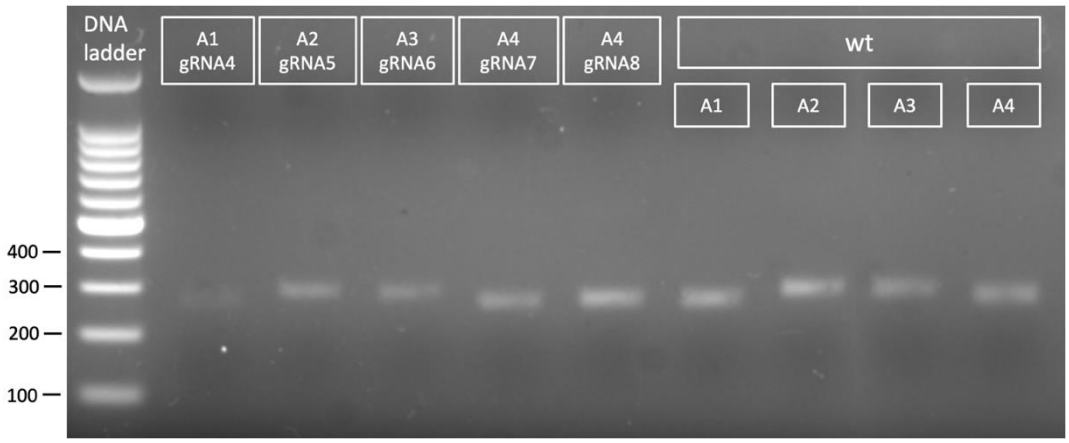


Figure 3.15 Gel electrophoresis of purified amplicons

1ul of purified DNA sample with 1ul of 6x DNA loading dye were loaded into each well on a 2% agarose gel, with a running time of 35mins at 100V. 5ul of 100bp DNA ladder was loaded in lane 1. Single band between 200 and 300 bp was detected for all amplicons generated from gRNA transfected cells and wt NIKS. No band due to primer dimer formation was detected. A = Amplicon.

Subsequently, gel electrophoresis was performed with the best optimised amplicon products generated from all five gRNAs (as highlighted in Figure 3.13) and wt NIKS (Figure 3.13 and 3.14). The single bands were extracted and purified, preventing the samples from containing any unspecific DNA molecules such as those due to primer dimer formation. Gel electrophoresis was repeated with 1ul of purified product (Figure 3.15) and an image was attached with the request form for targeted next generation sequencing (Illumina Miseq).

3.4.4 Next generation sequencing analysis

Gene editing efficiency of individual gRNAs was obtained from Miseq sequencing data (Figure 3.16). As indicated, gRNA 5 – 8 did not generate high frequency of indel mutations, which resulted in high wild type reads from the edited cell population. In contrast, gRNA 4 demonstrated a high percentage of *SYCP2* indels. However, these mutations occurred within a region that is upstream of the gRNA cut site where it contains a long tract of T nucleotides (Figure 3.17).

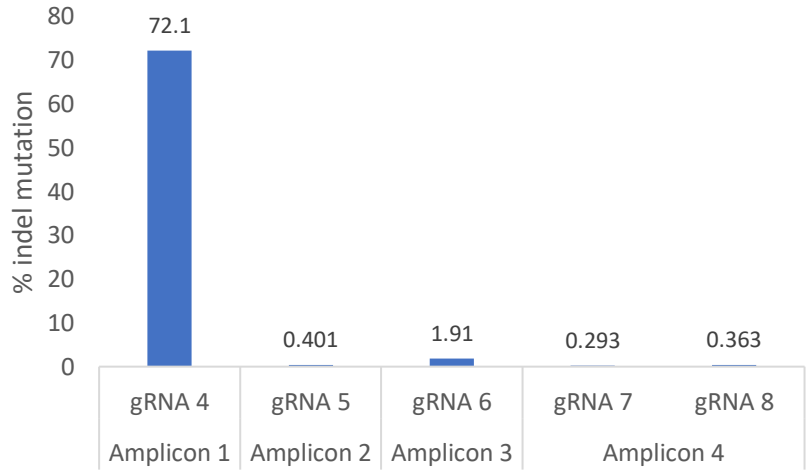


Figure 3.16 Gene editing efficiency of gRNAs

Results from Miseq sequencing analysis are presented as percentages of indel mutation due to delivery of CRISPR-Cas9 system to NIKS cells. gRNA 4 seems to demonstrate the most efficient gene editing ability, however the resulting indel mutations were not found at the gRNA target site.

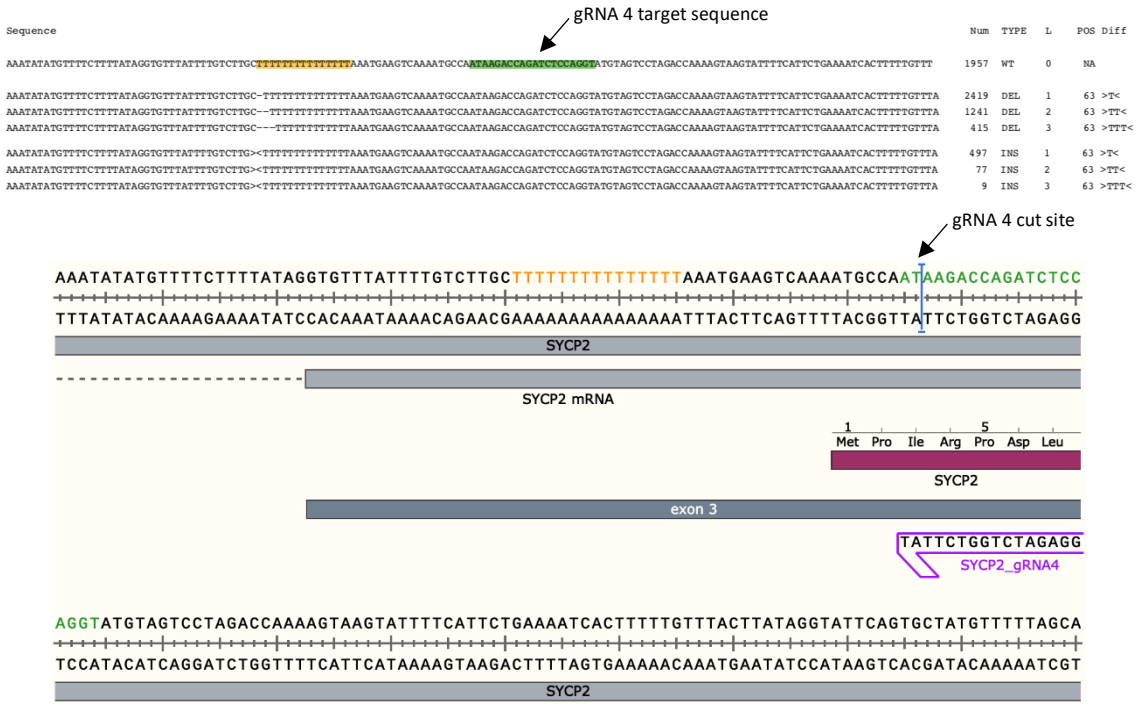


Figure 3.17 Illustration of the most common indel mutations generated by gRNA 4 and its corresponding target sequence and cut site

Three of the most common deletions and insertions occurred at 63bp where the repeating T nucleotides are found. The mutated site is upstream of the gRNA 4 target sequence as shown in SnapGene. No indels were observed at the gRNA cut site.

Miseq sequencing data for gRNA 4, 5, 7 and 8 showed poor gene editing efficiency with mutations occurring at non-gRNA cut site, gRNA 4 is shown as an example in Figure 3.17. Therefore, these data will not be further investigated in this project. Only gRNA 6 showed potential gene editing within the gRNA cut site. 261 deletion events occurred at 110bp with a single deleted G nucleotide, as illustrated in Figure 3.18. Wt NIKS cells did not display indel mutations at the gRNA 6 cute site (Figure 3.19), thereby indicating successful CRISPR activity. However, the cell population overwhelmingly consisted of wild type (17,294 reads), suggesting poor editing efficiency. Hence, single cell cloning was not carried out.

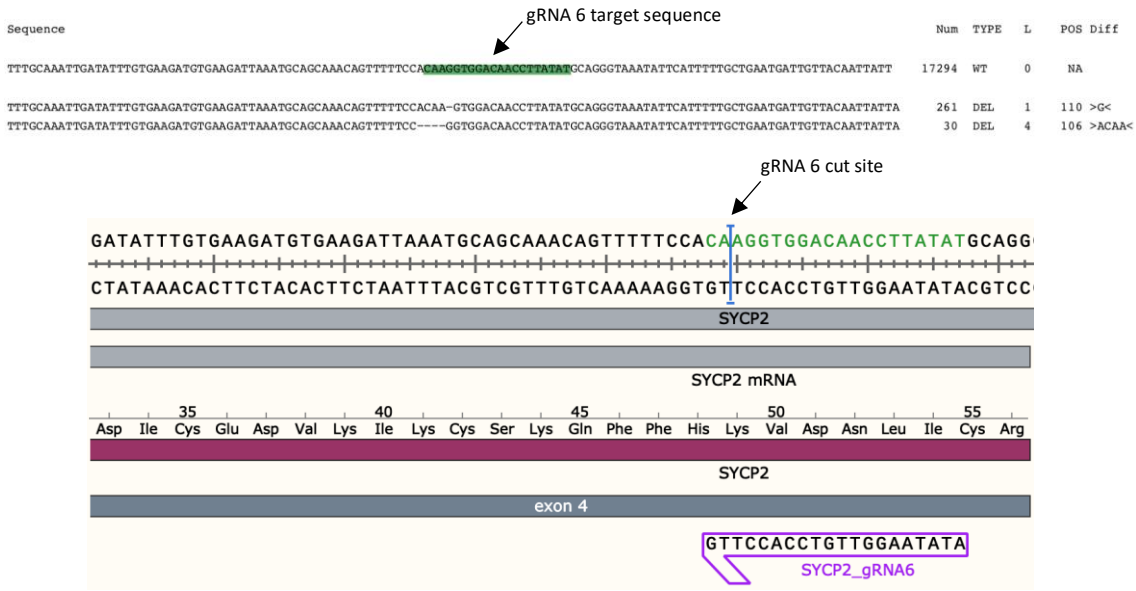


Figure 3.18 Illustration of the most common deletions generated by gRNA 6 and its corresponding target sequence and cut site

Most frequent deletions (total of 291 events) occurred around gRNA 6 cut site as indicated o SnapGene software. Including 261 single nucleotide (G) deletions and 30 of 4-nucleotide (ACAA) deletions.



Figure 3.19 Miseq sequencing data of amplicon 3 for wt NIKS cells

No indel mutations were generated at the gRNA 6 cut site for wt NIKS cells that was not transfected with gRNA 6.

3.4.5 CRISPR analysis

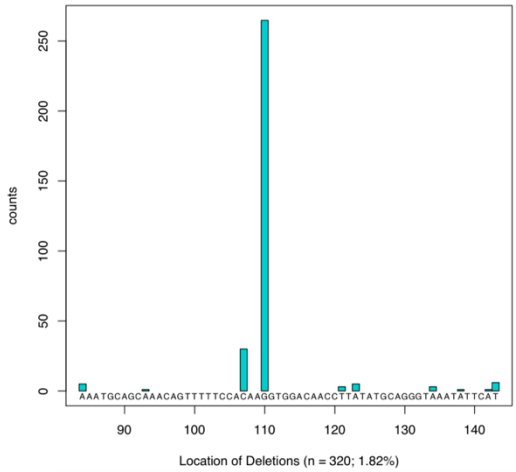


Figure 3.20 Graphic presentation of the most common deletions in cell population transfected with gRNA 6



Figure 3.21 The consequence of single nucleotide deletion within *SYCP2* gene

Single nucleotide deletion seen with gRNA 6 was revealed in SnapGene to induce a frameshift mutation immediate downstream of the cut site, causing incorrect translation of the genetic code. The new amino acid sequence is shown in the green box. Highlighted G is the deleted nucleotide.

CRISPR analysis provided graphic information on the size, location, and midpoint of most common indel mutations generated. Due to poor gene editing efficiency, only the location of gRNA 6 induced deletion is presented (Figure 3.20). Sequence alignment of the single nucleotide deletion using SnapGene revealed a frameshift mutation, causing an alteration in the amino acid sequence, as illustrated in Figure 3.21.

4 Discussion

4.1 SYCP2 mRNA is upregulated in HPV16+ cell lines

SYCP2 is a protein predominant found in the testes that is known to localise within the SC structure and contribute towards the organisation of chromosomal axis during meiotic prophase I. The ectopic expression of SYCP2 in HPV+ cancers has consequently led to the interest of this project, with the aim to elucidate whether knockout of SYCP2 expression could influence cancer cell viability.

The first objective was to validate the findings established in several HPV-related cancer studies, where significant *SYCP2* upregulation was shown. qPCR data collected from this project demonstrated a substantial increase of relative SYCP2 expression level in HPV16+ cancer cell lines in comparison to HPV- cell lines. Almost 30-fold increase was seen in 93-VU-147T and 20-fold with CaSki and SiHa compared to the HPV- cell lines, PCI-30 and C33a (Figure 3.5a). The HPV- cell lines showed almost no expression of SYCP2. In a study (Martinez *et al.* 2007) carried out with squamous cell carcinoma of the head and neck (SCCHN) patients identified *SYCP2* as one of the upregulated genes in both HPV+ versus HPV- SCCHN and HPV+ SCCHN versus normal oral tissues analysis. Together with data from Figure 3.4a suggests a strong correlation between HPV status and SYCP2 expression, however the mechanism is yet to be understood.

Interestingly, HeLa (HPV18+) showed little or no SYCP2 expression, corresponding to the unpublished RNAseq and DNA methylation data from a study (Fenton. T) where high SYCP2 levels were only detected in HPV16+ cervical cancer primary samples. This is a

possible indication for the existence of some key cellular differences between the genotypes. One of which could be the ability to upregulate SYCP2 expression.

Deregulated E6 and E7 expression are the fundamental events that drive malignancy in high risk HPV infections. Mainly through inactivation of tumour suppressor proteins and disruption of cell cycle progression (Xu *et al.* 2013). Figure 3.5b shows that all HPV16+ cell lines exhibit high E6 expression. In particular, CaSki showed over 500-fold increase in E6 copy number relative to TBP compared to HPV- cell lines. In contrast, SiHa only revealed approximately 50-fold increase in E6 copy number (Meissner 1999; Sahab *et al.* 2012;Xu *et al.* 2013). These data correspond to the HPV16 status of each cell line, CaSki has about 600 HPV16 copies per cell whereas SiHa only has 2 integrated copies of HPV16 genome. No apparent correlation can be drawn between the expression of E6 and SYCP2 in HPV16+ cell lines. For instance, CaSki demonstrated highest E6 level compared to other three cell lines but it is not resembled in the SYCP2 data. Furthermore, SiHa displayed similar SYCP2 level as CaSki despite being the cell line with the lowest E6 expression. However, it could be the case that a threshold level of E6 is required to induce SYCP2, thus SYCP2 expression plateaus once the threshold is reached. Expression of individual HPV proteins in NIKS could be performed to clarify if E6 expression has a direct effect on inducing SYCP2 expression.

4.2 Cell viability is not affected by *SYCP2* gene knockdown

The effect of *SYCP2* knockdown in HPV16+ cell lines did not seem to impact the cell viability as determined by MTS assay. In fact, higher percentage of viable cells were measured in *SYCP2*-silenced population compared to the culture transfected with NT siRNA (Figure 3.8a) which acts as a negative control. It should be noted that a significant decrease in cell viability was seen in NT relative to mock transfection. This indicate that the NT siRNA could be toxic, possibly due to off-target effects which caused cell death. Therefore, a different NT siRNA could be used in future work to prevent similar observation. Approximately 20% difference was observed between NT and SYCP2 data in all three HPV16+ cell lines. However, in some replicates, the cell viability percentages between NT and *SYCP2* transfected population showed little or almost no differences. The variation in data is most likely caused by frequent occurrence of freeze-thaw cycle, leading to reduced siRNA efficiency. Both graphs in Figure 3.8 seems to suggest that silencing of *SYCP2* gene improves the growth and survival of HPV16+ cancer cells.

Microscope image of *SYCP2* transfected UM-SCC-104 culture on day 5 post transfection (Figure 3.6) showed high similarity with the mock transfected population. This cell line tends to grow collectively in colonies which is displayed in both mock and SYCP2 transfected culture. In contrast, no colonies were seen in *PLK1*-silenced population (Figure 3.6c2) and majority of cancer cells were deteriorated with visible damaged cell membrane. Low viable cell percentage for *PLK1*-transfected population in all cell lines (Figure 3.8a) was a clear indication that high transfection efficiency was achieved. Both qualitative and quantitative data obtained from siRNA transfection experiments appears

to imply that the knockdown of SYCP2 expression does not have a negative effect on cancer cell viability. In fact, enhanced growth and survival seems to be more of the case.

Same concentration of siRNA was used in both quantification assays (MTS and qPCR) to allow direct analysis of gene knockdown for each cell line (Figure 3.9). The ON-TARGET siRNAs from Horizon were advertised to provide successful gene silencing at 75% or above. However, only about 50% gene knockdown was achieved in all tested cell lines. They still expressed SYCP2 at a level higher than HPV- cell lines (Figure 3.10). This reduction of mRNA level may not be sufficient to fully conclude the effect of SYCP2 knockdown on HPV+ cancer cell viability. Transfection efficiency was not a limiting factor as PLK1 data in Figure 3.8a demonstrated strikingly high knockdown percentages. Frequent freeze-thaw cycle could again cause this poor knockdown of *SYCP2* gene. However, it is unlikely to be the main contributing factor because similar knockdown efficiency was observed in experiments performed using new aliquots of *SYCP2* siRNA.

4.3 Possible mechanism for SYCP2 expression in HPV+ cancers

4.3.1 Epigenetic regulation

The precise mechanism and function of CTA expression in cancer remains poorly understood, but there have been many proposed theories such as epigenetic abnormality. A previous study (Litvinov *et al.* 2014) documented aberrant CTA expression in cutaneous T-cell lymphoma patients and discovered that inhibition of histone deacetylation resulted in significant upregulation of CTA mRNA level, including *SYCP1.* Cells treated with histone acetyltransferases inhibitors produced the opposite effect in which *SYCP1* was downregulated. Interestingly, histone deacetylase inhibitor

treatment resulted in downregulation of SYCP1 at the protein level. It was highlighted that many meiotic genes tend to undergo extensive post-translational regulation to ensure stability and facilitating different stages of spermatogenesis (Bettegowda and Wilkinson 2010). These combined results demonstrate a potential mode of regulation for CTA expression. Unpublished data from Tim Fenton observed a strong reduction in the methylation of multiple CpG sites in the SYCP2 promoter in HPV16+ but not HPV18+ or HPV25+ cervical cancers. Therefore, suggesting that the methylation status of SYCP2 promoter plays a potential role in inducing expression in HPV16+ cervical cancers.

Demethylation is another epigenetic event that regulate the expression of many genes, including *INK4A–ARF* which expresses a tumour suppressor protein named $p16^{INK4A}$. It is a critical regulator for cellular senescence. This gene is normally repressed in healthy cells by the presence of H3K27me3 (histone H3 lysine 27 trimethyl) mark, however it becomes activated upon cellular stress (Agger *et al.* 2009; McLaughlin-Drubin, Crum and Münger 2011). HPV16 E7 oncoprotein has been demonstrated to associate with a member of the E2F transcription factor family, E2F6 to induce the expression of two histone demethylases, KDM6A and KDM6B. E2F6 is a transcription repressor and a component of the polycomb transcriptional repressive complexes that associate with silenced chromatin (McLaughlin-Drubin, Huh and Münger 2008). KDM6B is responsible for the demethylation of H3K27m3 in HPV+ cancers that result in the expression of $p16^{INK4A}$. High $p16^{INK4A}$ level is frequently found in HPV-related cancers such as cervical carcinomas. Thus, it is recognised as a biomarker for cervical cancers. Furthermore, a significant reduction of H3K27m3 mark was shown in primary human foreskin keratinocyte (HFK) with HPV16 E7 expression. Whereas no detectable changes were observed in the control keratinocytes (McLaughlin-Drubin, Crum and Münger 2011).

A study (McLaughlin-Drubin, Park and Munger 2013) performed with HPV16+ cell lines including CaSki and SiHa demonstrated a dramatic decrease in cell proliferation/survival when KDM6B level was depleted. Ectopic expression of HPV16 E7 in HFK with depleted KDM6B also showed a striking decrease in cell viability. Hence, it was concluded that the KDM6B addiction in cervical cancer cells was a direct consequence of HPV16 E7 expression. Therefore, demethylation of *SYCP2* promoter via this mechanism could potentially be responsible for the aberrant expression in HPV16+ cancers. However further investigations will be required to elucidate this theory.

4.3.2 Genomic instability

As mentioned previously, SYCP2 is an important component of the SC structure which ensures high-fidelity meiotic segregation of homologous chromosome during prophase I. Ectopic expression of SYCP2 in HPV-positive cancers including OPSCC, HNSCC and cervical squamous cell carcinoma (CSCC) likely contribute to genomic instability which is a hallmark of cancer (Guo *et al.* 2014; L. Costa, Boroni and Soares 2017). This characteristic allows cancer progression and prevent immunologic and pharmacologic eradication.

A review (Lindsey *et al.* 2013) described possible consequences of meiotic gene expression in mitotic cells. In particular, *SYCP1* was one of the genes that were expressed in melanomas and it may contribute to chromosomal missegregation and genomic instability. The review particularly focused on other meiotic proteins such as SPO11 and REC8. REC8 is a meiotic-specific cohesin protein that replaces mitotic protein RAD21. This protein serves many function including cooperation with SC to promote chiasma

formation, homologous recombination, and adhesion between sister chromatids. Overexpression of REC8 in mitotic cells has been noted to cause chromosome segregation defects (Ishiguro *et al.* 2010).

Studies have shown that SYCP3 expression in mitotic cells seems to inhibit an important component of the homologous recombination (HR) repair pathway, BRCA2. In addition, it caused cells to confer hypersensitivity to poly(ADP-ribose) polymerase (PARP) inhibitors (Hosoya *et al.* 2011; Hosoya and Miyagawa 2014). This characteristic is commonly observed in cancer therapy with PARP inhibitors for BRCA1/BRCA2 deficient tumours. Generally, cells respond to DNA damage such as DSBs by activating BRCA2 (a tumour suppressor protein) which recruits a protein called RAD51 to the DSB site, where it is loaded onto single-stranded DNA to promote homologous DNA paring. However, when SYCP3 is expressed, the HR pathway is disrupted and consequently causing DSB accumulation which ultimately leads to genomic instability. This finding provided a potential mechanism for chromosomal instability seen in cancers with SYCP3 overexpression. Thereby it would be interesting to investigate if the HR pathway is also impaired/altered in HPV+ cancers with aberrant SYCP2 expression.

A study (Li, Bolcun-Filas and Schimenti 2011) demonstrated that *SYCP3*-decifient female mice produced viable oocytes but often with aneuploidy features. It was discovered that the expression of DNA meiotic recombinase 1 (DMC1) and TRIP13, proteins normally required for HR repair of meiotic DSBs, is deemed non-essential for oocyte survival in the absence of SYCP3 and SYCP2. This bypass involves a protein, DNA repair and recombination protein RAD54 (RAD54) that stimulates inter-sister chromatid recombination in mammalian mitotic cells and yeast meiosis. Essentially, the oocytes

arrest if inter-homolog repair of DSBs were prevented, proving that normal oocytes can only repair meiotic DSBs via recombination between homologous chromosomes, and not via recombination between sister chromatids. However, if SYCP2 or SYCP3 expression is disrupted, then the oocytes survived. It was concluded that the surviving oocytes were able to repair their DSBs via inter-sister recombination. Therefore, SYCP2, SYCP3 or both must act to prevent inter-sister recombination and repair. This could explain why HPV16 induce SYCP2 expression, inter-sister recombinational repair of the host genome would be prevented and thus maintaining the G2 arrest state, allowing viral amplification.

HPV E6 and E7 expression was established in a study (Wallace *et al.* 2017)to hinder the HR repair pathway. It promoted the initiation of HR at particular cell cycle stages where the sister chromatids were unavailable to serve as a homologous template. Furthermore, E6 was capable of preventing RAD51 recruitment to both persistent and transient DSB sites. Consequently, the inability to repair DSBs would cause persistent lesions in cancers, which is a common phenotype observed in tumours with increased integration of HPV genome. However, it was also demonstrated that both HPV16 oncogenes stimulated expression of RAD51 and BRCA2 which are components of the HR pathway. Therefore, it was suggested that E6 and E7 promote initiation but inhibit the full completion of HR. This was based on previous findings showing increased level of RAD51 localisation at the HPV replication foci upon genome amplification (Gillespie *et al.* 2012), implicating the need for HR protein during the viral productive phase. Although full investigation has not been previously carried out on SYCP2 overexpression, the research on meiotic-specific genes such as SYCP3 may provide a fundamental platform for uncovering possible mechanisms responsible for aberrant SYCP2 expression in HPV+ cancers.

4.4 Application of CRISPR-Cas9 system in NIKS cells

CRISPR-Cas9 system is a highly feasible and flexible gene editing technology, with great specificity at the DNA level. Allowing targeting of given sequences at any locus. Hence it has been widely used in cancer biology to screen for any cancer-related genes and establishing protein function through knockout mutations (Wang *et al.* 2014). Complete *SYCP2* knockdown in NIKS cells using CRISPR were subsequently performed after insufficient knockdown percentage in siRNA experiments. NIKS cells were used in this project for their non-tumorigenic feature whilst maintaining normal cell-type-specific growth in monolayer culture (Allen-Hoffmann *et al.* 2000). They also exhibit limitless capacity to proliferate due to their immortalised state.

A study (Yamano, Dai and Moursi 2010) has reported on the transfection efficiency of FuGene HD reagent in different cell lines, including several cancer cells lines, stem cells and primary human epidermal keratinocytes. Only 2-5% efficiency was achieved in transfection of β-Galactosidase to primary human epidermal keratinocytes. Whereas, 20-35% was seen in HeLa cell line. Overall, FuGene HD transfection reagent was proven to exhibit high transfection efficiency.

Five set of gRNAs were designed to contain sequences complementary to exon 3, 4 and 5. The aim was to generate random indel mutations and cause disruption in SYCP2 structure and function through the NHEJ DNA repair pathway. Followed by generating single cell clones with desired indel mutations. However, none of the gRNA successfully induced indel mutation at the gRNA cut site with sufficient editing efficiency. Only gRNA 6 showed 291 deletion events at the relevant site (Figure 3.18), but <2% efficiency was

accomplished with over 17,000 wt read. Therefore, single cell cloning was not proceeded due to the process being time-consuming with little chance of success.

Chromatin status has been found to relate to CRISPR transfection efficiency. The prokaryotic origin of Cas9 nuclease may prevent its full operation in different eukaryotic chromatin structures due to the fundamental differences. Evidence has shown that longer scanning time is required when the target site is hidden within heterochromatin compared to euchromatin where the site is more accessible by Cas9 (Uusi-Mäkelä *et al.* 2018; Verkuijl and Rots 2019). In addition, a study (Hinz, Laughery and Wyrick 2015a) has shown that the catalytic activity of Cas9 significantly decreased when the PAM motif was localised within the nucleosome compared to if it was situated in the more accessible DNA linker region. Hence, it was speculated that more inhibitory effect will be displayed at higher-order chromatin structure. Findings have suggested that off-site target binding strongly correlates with the accessibility of chromatin and gRNAs are known to vary widely in their off-target binding frequency (Kuscu *et al.* 2014; Hinz, Laughery and Wyrick 2015b). These combined results demonstrated the positive correlation between open chromatin state and gRNA targeting efficiency. Hence, it is possible that the *SYCP2* gRNA target sites were buried within the closed chromatin in wt NIKS, which impeded the accessibility of Cas9 to its relevant DNA cut site.

Flow cytometry was performed post transfection to produce a sorted population containing GFP-positive cells. The FACS data (Figure 3.12, bottom panels) showed a peak towards the end of the plot for all gRNA transfected cells, which is not seen in the wt NIKS population. Therefore, it was speculated that the voltage was set too high during sorting, leading to the desired GFP-positive cells not been included within the gated

population. This indicates that the cells were sorted due to autofluorescence instead of GFP signal. Hence, this would explain the observed data where the sorted population contained primarily of wt NIKS for all gRNAs.

CRISPR transfection was attempted again but due to unavailability of the equipment, the post transfection cell samples remained unsorted. Regardless, DNA was still extracted with the intention to send the unsorted samples for NGS analysis. Due to the time constraints of this project, the NGS report will not be incorporated in this project, however it would allow the project to proceed in the future by providing some insight into which gRNA would display the highest editing efficiency.

4.5 Established quantification assay for *SYCP2* mRNA

This project successfully established a novel quantification assay for SYCP2 at the mRNA level. 100% amplification efficiency is achieved when the amount of molecule of the target sequence doubles during each amplification cycle. The SYCP2 standard curve generated displayed 99% primer efficiency, indicating appropriate melting temperature and great primer design with no formation of secondary structures such as dimers and hairpins. Therefore, it was redeemed acceptable to apply the SYCP2 standard curve in qPCR analysis.

Other analysis method includes the ΔΔCT tool which can be used to compare target gene expression in relation to other selected genes. The data obtained is presented in fold changes. This method was initially used in this project, where SYCP2 expression in HPV+ cell lines was displayed as the number of fold changes it is increased by when compared to HPV- cell lines. However, the collected data showed significant variation between repeated experiments. Therefore, the standard curve was generated as a new quantification assay. It is a particularly useful tool that allows direct comparison of SYCP2 expression with normalisation to housekeeping genes such as *TBP* between any cell lines or primary cells. It also allowed us to check the efficiency of SYCP2 primers. The process may be time-consuming, but the standard curve can be used in any future work and allows direct comparison with the results collected from this project. Furthermore, the procedure described in section 2.5 can be adapted to generate standard curve for any other gene of interest.

4.6 Conclusion and future directions

4.6.1 Conclusion

To recapitulate, whilst protein expression was unable to be established in HPV16+ cell lines, SYCP2 was shown to be overexpressed at the mRNA level, which was quantified using the standard curve generated. This corresponds to the studies on HPV+ cancers where *SYCP2* was identified as one of the genes that were upregulated in relation to HPV- cancers (Pyeon *et al.* 2007; Schlecht *et al.* 2007; Masterson *et al.* 2015). The inability to induce complete knockdown in siRNA transfection despite high PLK1 efficiency raised an interesting assumption.

Unfortunately, CRISPR knockout in NIKS cells were not successful, possibly due to the inaccessibility of chromatin structure or high voltage during cell sorting. Nonetheless, some level of CRISPR knockout was seen with gRNA 6, where indel mutations were detected within the target cut site, causing frameshift mutation with altered amino acid sequence. This could possibly hinder the cellular function of SYCP2. However, low editing efficiency prevented further experiments such as generating single cell clones. In conclusion, with better understanding and clarification of the mechanism responsible for aberrant SYCP2 expression in mitotic cells and cancer cells, it would allow the development of potential novel biomarkers for several HPV-induced cancers and be utilised for immunotherapy treatment.

4.6.2 Future directions

If SYCP2 is deemed as critical for cellular survival, then it would be possible that cells with silenced *SYCP2* gene would display reduced viability. Hence it would be interesting to attempt knockout in HPV+ cell lines such as SiHa and UM-SCC-104 with known SYCP2 expression. Followed by direct comparison with the qPCR data collected from this project to show if SYCP2 depletion has any negative effect on cancer cell survival.

SYCP2 protein expression has not yet been fully confirmed in previous study (Pyeon *et al.* 2007), possibly due to protein instability which leads to intrinsic degradation. Treating the HPV+ cancer cells with membrane permeable proteasome inhibitors such as MG132 could resolve the degradation issue and confirm the protein expression using Western blot analysis (Yong *et al.* 2009).

Previous study demonstrated that E7 knockout in NIKS with HPV16 showed partially reduced SYCP2 expression when compared to NIKS with HPV16 (Pyeon *et al.* 2007). This indicated the induction effect of HPV16 E7 on SYCP2. Therefore, E6 knockout could also be tested in HPV+ cells lines to see if it demonstrates similar effect. In addition, combined knockout would allow the confirmation of synergistic effect of E6 and E7 induction on SYCP2 expression. Upon obtaining knockout models, E6/E7 expression can be re-introduced to see if SYCP2 expression returns to the level detected in HPV16+ cancer cell lines.

5. References

Adhikary, S. and Eilers, M. (2005). Transcriptional regulation and transformation by Myc proteins. *Nature Reviews Molecular Cell Biology* **6**:635–645.

Agger, K. *et al.* (2009). The H3K27me3 demethylase JMJD3 contributes to the activation of the INK4A-ARF locus in response to oncogene- and stress-induced senescence. *Genes & development* **23**:1171–1176.

Agrawal, N. *et al.* (2003). RNA Interference: Biology, Mechanism, and Applications. *Microbiology and Molecular Biology Reviews* **67**:657 LP – 685.

Aksoy, P., Gottschalk, E.Y. and Meneses, P.I. (2017). HPV entry into cells. *Mutation Research* **772**:13–22.

Allen-Hoffmann, B.L. *et al.* (2000). Normal Growth and Differentiation in a Spontaneously Immortalized Near-Diploid Human Keratinocyte Cell Line, NIKS. *Journal of Investigative Dermatology* **114**:444–455.

Alves, I. *et al.* (2017). The impact of recombination on human mutation load and disease. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**.

Aydin, I. *et al.* (2014). Large Scale RNAi Reveals the Requirement of Nuclear Envelope Breakdown for Nuclear Import of Human Papillomaviruses. *PLoS Pathogens* **10**.

Banerjee, N.S. *et al.* (2011). Human papillomavirus (HPV) E7 induces prolonged G 2 following S phase reentry in differentiated human keratinocytes. *Journal of Biological Chemistry* **286**:15473–15482.

Bettegowda, A. and Wilkinson, M.F. (2010). Transcription and post-transcriptional regulation of spermatogenesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **365**:1637–1651.

Bienkowska-Haba, M. *et al.* (2018). A new cell culture model to genetically dissect the complete human papillomavirus life cycle. *PLoS Pathogens* **14**:1–21.

Blackford, A.N. and Jackson, S.P. (2017). ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. *Molecular Cell* **66**:801–817.

Bode, P.K. *et al.* (2014). Cancer testis antigen expression in testicular germ cell tumorigenesis. *Modern Pathology* **27**:899–905.

Bravo, I.G. and Felez-Sanchez, M. (2015). Papillomaviruses: Viral evolution, cancer and evolutionary medicine. *Evolution, Medicine and Public Health* **2015**:32–51.

Bravo, I.G., de Sanjosé, S. and Gottschling, M. (2010). The clinical importance of understanding the evolution of papillomaviruses. *Trends in Microbiology* **18**:432–438.

Bray, F. *et al.* (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68**:394–424.

Broniarczyk, J. *et al.* (2018). HPV-16 virions can remain infectious for 2 weeks on senescent cells but require cell cycle re-activation to allow virus entry. *Scientific Reports* **8**:1–11.

Bruggeman, J.W. *et al.* (2018). Massive expression of germ cell-specific genes is a hallmark of cancer and a potential target for novel treatment development. *Oncogene* **37**:5694–5700.

Bruni, L. *et al.* (2010). Cervical Human Papillomavirus Prevalence in 5 Continents: Meta-Analysis of 1 Million Women with Normal Cytological Findings. *The Journal of Infectious Diseases* **202**:1789–1799.

Buck, C., Day, P. and Trus, B. (2008). The papiloma Major Capsid Protein L1. *Virology* **23**:1–7.

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**:4132–4141.

Ciró, M. *et al.* (2009). ATAD2 is a novel cofactor for MYC, overexpressed and amplified in aggressive tumors. *Cancer Research* **69**:8491–8498.

Costa, J.R. *et al.* (2004). Genome Editing Using Engineered Nucleases and Their Use in Genomic Screening. *Assay Guidance Manual*:1–24.

Culp, T.D. *et al.* (2006). Keratinocyte-Secreted Laminin 5 Can Function as a Transient Receptor for Human Papillomaviruses by Binding Virions and Transferring Them to Adjacent Cells. *Journal of Virology* **80**:8940–8950.

Danielewski, J.A. *et al.* (2013). Human Papillomavirus Type 6 and 11 Genetic Variants Found in 71 Oral and Anogenital Epithelial Samples from Australia. *PLoS ONE* **8**:1–9.

Day, P.M., Lowy, D.R. and Schiller, J.T. (2008). Heparan Sulfate-Independent Cell Binding and Infection with Furin-Precleaved Papillomavirus Capsids. *Journal of Virology* **82**:12565–12568.

Depuydt, C.E. *et al.* (2016). Human Papillomavirus (HPV) virion induced cancer and subfertility, two sides of the same coin. *Facts, views & vision in ObGyn* **8**:211–222.

Dobrynin, P. *et al.* (2013). The novelty of human cancer/testis antigen encoding genes in evolution. *International Journal of Genomics* **2013**.

Di Domenico, F. *et al.* (2009). Expression of human papilloma virus type 16 E5 protein in amelanotic melanoma cells regulates endo-cellular pH and restores tyrosinase activity. *Journal of Experimental and Clinical Cancer Research* **28**:99–114.

Doorbar, J. *et al.* (2012). The biology and life-cycle of human papillomaviruses. *Vaccine* **30**:F55–F70.

Doorbar, J. (2013). The E4 protein; structure, function and patterns of expression. *Virology* **445**:80–98.

Van Doorslaer, K. *et al.* (2018). ICTV virus taxonomy profile: Papillomaviridae. *Journal of General Virology* **99**:989–990.

Dreer, M., van de Poel, S. and Stubenrauch, F. (2017). Control of viral replication and transcription by the papillomavirus E8^E2 protein. *Virus Research* **231**:96–102.

Egawa, N. *et al.* (2017). HPV16 and 18 genome amplification show different E4-dependence, with 16E4 enhancing E1 nuclear accumulation and replicative efficiency via its cell cycle arrest and kinase activation functions. *PLoS Pathogens* **13**.

Epping, M.T. *et al.* (2005). The human tumor antigen PRAME is a dominant repressor of retinoic acid receptor signaling. *Cell* **122**:835–847.

Feng, J. *et al.* (2017). Synaptonemal complex protein 2 (SYCP2) mediates the association of the centromere with the synaptonemal complex. *Protein & cell* **8**:538–543.

Fire, A. *et al.* (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**:806–811.

Formana, D. *et al.* (2012). Global burden of human papillomavirus and related diseases. *Vaccine* **30**:F12–F23.

Francis, D.A., Schmid, S.I. and Howley, P.M. (2000). Repression of the Integrated Papillomavirus E6/E7 Promoter Is Required for Growth Suppression of Cervical Cancer Cells. *Journal of Virology* **74**:2679–2686.

Fratta, E. *et al.* (2011). The biology of cancer testis antigens: Putative function, regulation and therapeutic potential. *Molecular Oncology* **5**:164–182.

Fraune, J. *et al.* (2012). The mammalian synaptonemal complex: Protein components, assembly and role in meiotic recombination. *Experimental Cell Research* **318**:1340–1346.

Gillespie, K.A. *et al.* (2012). Human Papillomaviruses Recruit Cellular DNA Repair and Homologous Recombination Factors to Viral Replication Centers. *Journal of Virology* **86**:9520 LP – 9526.

Gjerstorff, M.F., Andersen, M.H. and Ditzel, H.J. (2015). Oncogenic cancer/testis antigens: Prime candidates for immunotherapy. *Oncotarget* **6**:15772–15787.

Graham, S. V. (2010). Human papillomavirus: Gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. *Future Microbiology* **5**:1493–1506.

Graham, S. V. (2017). The human papillomavirus replication cycle, and its links to cancer progression: A comprehensive review. *Clinical Science* **131**:2201–2221.

Guan, P. *et al.* (2012). Human papillomavirus types in 115,789 HPV-positive women: A meta-analysis from cervical infection to cancer. *International Journal of Cancer* **131**:2349–2359.

Guo, P. *et al.* (2014). The landscape of alternative splicing in cervical squamous cell carcinoma. *OncoTargets and therapy* **8**:73–79.

Handisurya, A., Schellenbacher, C. and Kirnbauer, R. (2009). Diseases caused by human papillomaviruses (HPV). *Journal der Deutschen Dermatologischen Gesellschaft* **7**:453–466.

Harari, A., Chen, Z. and Burk, R.D. (2014). Human papillomavirus genomics: Past, present and future. *Current Problems in Dermatology (Switzerland)* **45**:1–18.

Hinz, J.M., Laughery, M.F. and Wyrick, J.J. (2015a). Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro. *Biochemistry* **54**:7063–7066.

Horvath, C.A. *et al.* (2010). Mechanisms of cell entry by human papillomaviruses: An overview. *Virology Journal* **7**:1–7.

Hosoya, N. *et al.* (2011). Synaptonemal complex protein SYCP3 impairs mitotic recombination by interfering with BRCA2. *EMBO reports* **13**:44–51.

Hosoya, N. and Miyagawa, K. (2014). Targeting DNA damage response in cancer therapy. *Cancer science* **105**:370–388.

Ishiguro, T. *et al.* (2010). Shugoshin–PP2A counteracts casein-kinase-1-dependent cleavage of Rec8 by separase. *Nature Cell Biology* **12**:500–506.

Jiang, P. and Yue, Y. (2013). Human papillomavirus oncoproteins and apoptosis (Review). *Experimental and Therapeutic Medicine* **7**:3–7.

Jinek, M. *et al.* (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)* **337**:816–821.

Jinek, M. *et al.* (2014). Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* **343**:1247997.

Jing, K. *et al.* (2014). Docosahexaenoic acid induces the degradation of HPV E6/E7 oncoproteins by activating the ubiquitin-proteasome system. *Cell Death and Disease* **5**:1–11.

Kajitani, N. *et al.* (2012). Productive lifecycle of human papillomaviruses that depends upon squamous epithelial differentiation. *Frontiers in Microbiology* **3**:1–12.

Kalashnikova, E. V. *et al.* (2010). ANCCA/ATAD2 overexpression identifies breast cancer patients with poor prognosis, acting to drive proliferation and survival of triple-negative cells through control of B-Myb and EZH2. *Cancer Research* **70**:9402–9412.

Karpf, A.R. (2006). Potential role for epigenetic modulatory drugs in the enhancement of cancer/germ-line antigen vaccine efficacy. *Epigenetics* **1**:116–120.

Kouznetsova, A. *et al.* (2005). SYCP2 and SYCP3 are required for cohesin core integrity at diplotene but not for centromere cohesion at the first meiotic division. *Journal of Cell Science* **118**:2271 LP – 2278.

Kuscu, C. *et al.* (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnology* **32**:677.

L. Costa, R., Boroni, M. and Soares, M.A. (2017). Distinct co-expression networks using multi-omic data reveal novel interventional targets in HPV-positive and negative head-and-neck squamous cell cancer. *bioRxiv*:236133.

Lam, J.K.W. *et al.* (2015). siRNA Versus miRNA as Therapeutics for Gene Silencing. *Molecular Therapy - Nucleic Acids* **4**:e252.

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**:71–82.

Lindsey, S.F. *et al.* (2013). Potential Role of Meiosis Proteins in Melanoma Chromosomal Instability. *Journal of Skin Cancer* **2013**:1–9.

Litvinov, I. V *et al.* (2014). Ectopic expression of cancer-testis antigens in cutaneous T-cell lymphoma patients. *Clinical cancer research : an official journal of the American Association for Cancer Research* **20**:3799–3808.

Liu, Y. *et al.* (2017). Whole-genome analysis of human papillomavirus types 16, 18, and 58 isolated from cervical precancer and cancer samples in Chinese women. *Scientific Reports* **7**:1–9.

Martinez, I. *et al.* (2007). Identification of differentially expressed genes in HPV-positive and HPV-negative oropharyngeal squamous cell carcinomas. *European journal of cancer (Oxford, England : 1990)* **43**:415–432.

Masterson, L. *et al.* (2015). Deregulation of SYCP2 predicts early stage human papillomavirus-positive oropharyngeal carcinoma: A prospective whole transcriptome analysis. *Cancer science* **106**:1568–1575.

McBride, A.A. and Warburton, A. (2017). The role of integration in oncogenic progression of HPV-associated cancers. *PLoS Pathogens* **13**:1–7.

McLaughlin-Drubin, M.E., Crum, C.P. and Münger, K. (2011). Human papillomavirus E7 oncoprotein induces KDM6A and KDM6B histone demethylase expression and causes epigenetic reprogramming. *Proceedings of the National Academy of Sciences* **108**:2130 LP – 2135.

McLaughlin-Drubin, M.E., Huh, K.-W. and Münger, K. (2008). Human Papillomavirus Type 16 E7 Oncoprotein Associates with E2F6. *Journal of Virology* **82**:8695 LP – 8705.

McLaughlin-Drubin, M.E., Park, D. and Munger, K. (2013). Tumor suppressor $p16^{INK4A}$ is necessary for survival of cervical carcinoma cell lines. *Proceedings of the National Academy of Sciences* **110**:16175 LP – 16180.

Meissner, J.D. (1999). Nucleotide sequences and further characterization of human papillomavirus DNA present in the CaSki, SiHa and HeLa cervical carcinoma cell lines. *Journal of General Virology* **80**:1725–1733.

Middleton, K. *et al.* (2003). Organization of Human Papillomavirus Productive Cycle during Neoplastic Progression Provides a Basis for Selection of Diagnostic Markers. *Journal of Virology* **77**:10186–10201.

Müller, M. *et al.* (2015). Human papillomavirus E5 oncoprotein: Function and potential target for antiviral therapeutics. *Future Virology* **10**:27–39.

Nakahara, T. *et al.* (2005). Human Papillomavirus Type 16 E1 E4 Contributes to Multiple Facets of the Papillomavirus Life Cycle. *Journal of Virology* **79**:13150–13165.

Nakahara, T. and Kiyono, T. (2016). Interplay between NF-κB/interferon signaling and the genome replication of HPV. *Future Virology* **11**:141–155.

Nees, M. *et al.* (2000). Human papillomavirus type 16 E6 and E7 proteins inhibit differentiation-dependent expression of transforming growth factor-β2 in cervical keratinocytes. *Cancer Research* **60**:4289–4298.

Nejo, Y.T., Olaleye, D.O. and Odaibo, G.N. (2018). Prevalence and Risk Factors for Genital Human Papillomavirus Infections Among Women in Southwest Nigeria. *Archives of basic and applied medicine* **6**:105–112.

Offenberg, H.H. *et al.* (1998). SCP2: A major protein component of the axial elements of synaptonemal complexes of the rat. *Nucleic Acids Research* **26**:2572–2579.

Oldak, M. *et al.* (2004). The Human Papillomavirus Type 8 E2 Protein Suppresses 4-Integrin Expression in Primary Human Keratinocytes. *Journal of Virology* **78**:10738–10746.

OMS (2007). *INTERNATIONAL AGENCY FOR RESEARCH ON CANCER IARC Monographs on the Evaluation of Carcinogenic Risks to Humans VOLUME 90 Human Papillomaviruses*. Vol. 80.

Orav, M. *et al.* (2015). Initial amplification of the HPV18 genome proceeds via two distinct replication mechanisms. *Scientific Reports* **5**:1–16.

Ozaki, T. and Nakagawara, A. (2011). Role of p53 in cell death and human cancers. *Cancers* **3**:994–1013.

Park, R.B. and Androphy, E.J. (2002). Genetic Analysis of High-Risk E6 in Episomal Maintenance of Human Papillomavirus Genomes in Primary Human Keratinocytes. *Journal of Virology* **76**:11359–11364.

Patel, D. *et al.* (2004). Human Papillomavirus Type 16 E6 and E7 Cause Polyploidy in Human Keratinocytes and Up-Regulation of G2-M-Phase Proteins. *Cancer Research* **64**:1299–1306.

Pellagatti, A. *et al.* (2015). Application of CRISPR/Cas9 genome editing to the study and treatment of disease. *Archives of Toxicology* **89**:1023–1034.

Pelttari, J. *et al.* (2001). A Meiotic Chromosomal Core Consisting of Cohesin Complex Proteins Recruits DNA Recombination Proteins and Promotes Synapsis in the Absence of an Axial Element in Mammalian Meiotic Cells. *Molecular and Cellular Biology* **21**:5667 LP – 5677.

Petrova, N.S., Zenkova, M.A. and Chernolovskaya, E.L. (2013). Structure - Functions Relations in Small Interfering RNAs. *Practical Applications in Biomedical Engineering*.

Pickar-Oliver, A. and Gersbach, C.A. (2019). The next generation of CRISPR–Cas technologies and applications. *Nature Reviews Molecular Cell Biology* **20**:490–507.

Pinidis, P. *et al.* (2016). Human Papilloma Virus' Life Cycle and Carcinogenesis. *Maedica* **11**:48–54.

Pirog, E.C. *et al.* (2014). HPV prevalence and genotypes in different histological subtypes of cervical adenocarcinoma, a worldwide analysis of 760 cases. *Modern Pathology* **27**:1559–1567.

Prieto, I. *et al.* (2001). Mammalian STAG3 is a cohesin specific to sister chromatid arms in meiosis I. *Nature Cell Biology* **3**:761–766.

Pyeon, D. *et al.* (2007). Fundamental Differences in Cell Cycle Deregulation in Human Papillomavirus–Positive and Human Papillomavirus–Negative Head/Neck and Cervical Cancers. *Cancer Research* **67**:4605 LP – 4619.

Reinson, T. *et al.* (2013). Engagement of the ATR-Dependent DNA Damage Response at the Human Papillomavirus 18 Replication Centers during the Initial Amplification. *Journal of Virology* **87**:951–964.

Reinson, T. *et al.* (2015). The cell cycle timing of human papillomavirus DNA replication. *PLoS ONE* **10**:1–16.

Richards, R.M. *et al.* (2006). Cleavage of the papillomavirus minor capsid protein, L2, at a furin consensus site is necessary for infection. *Proceedings of the National Academy of Sciences of the United States of America* **103**:1522–1527.

Sahab, Z. *et al.* (2012). Quantitative measurement of human papillomavirus type 16 e5 oncoprotein levels in epithelial cell lines by mass spectrometry. *Journal of virology* **86**:9465–9473.

Salman, N.A. *et al.* (2017). Association of high risk human papillomavirus and breast cancer: A UK based Study. *Scientific Reports* **7**:1–8.

Salmaninejad, A. *et al.* (2016). Cancer/Testis Antigens: Expression, Regulation, Tumor Invasion, and Use in Immunotherapy of Cancers. *Immunological Investigations* **45**:619–640.

Sapp, M.J. (2013). HPV virions hitchhike a ride on retromer complexes. *Proceedings of the National Academy of Sciences of the United States of America* **110**:7116–7117.

Schiffman, M. *et al.* (2016). Carcinogenic human papillomavirus infection. *Nature Reviews Disease Primers* **2**.

Schlecht, N.F. *et al.* (2007). Gene expression profiles in HPV-infected head and neck cancer. *The Journal of Pathology* **213**:283–293.

Seo, E.K. *et al.* (2016). Crystal structure of C-terminal coiled-coil domain of SYCP1 reveals non-canonical anti-parallel dimeric structure of transverse filament at the synaptonemal complex. *PLoS ONE* **11**:1–13.

Setten, R.L., Rossi, J.J. and Han, S. (2019). The current state and future directions of RNAi-based therapeutics. *Nature Reviews Drug Discovery* [Online] **18**:421–446.

Sigalotti, L. *et al.* (2002). Cancer testis antigens expression in mesothelioma: Role of DNA methylation and bioimmunotherapeutic implications. *British Journal of Cancer* **86**:979–982.

Sigalotti, Luca *et al.* (2002). Promoter Methylation Controls the Expression of MAGE2, 3 and 4 Genes in Human Cutaneous Melanoma. *Journal of Immunotherapy* **25**:16–26.

da Silva, V.L. *et al.* (2017). Genome-wide identification of cancer/testis genes and their association with prognosis in a pan-cancer analysis. *Oncotarget* **8**:92966–92977.

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**:701 LP – 709.

Smith, J.A. *et al.* (2014). SMCX and components of the TIP60 complex contribute to E2 regulation of the HPV E6/E7 promoter. *Virology* **468**:311–321.

Stanley, M.A. (2012). Epithelial cell responses to infection with human papillomavirus. *Clinical Microbiology Reviews* **25**:215–222.

Stoler, M.H. *et al.* (1992). Human papillomavirus type 16 and 18 gene expression in cervical neoplasias. *Human Pathology* **23**:117–128.

Surviladze, Z., Dziduszko, A. and Ozbun, M.A. (2012). Essential roles for soluble virion-associated heparan sulfonated proteoglycans and growth factors in human papillomavirus infections. *PLoS Pathogens* **8**.

Tomaić, V. (2016). Functional roles of E6 and E7 oncoproteins in HPV-induced malignancies at diverse anatomical sites. *Cancers* **8**.

Tuominen, H. *et al.* (2018). HPV infection and bacterial microbiota in the placenta, uterine cervix and oral mucosa. *Scientific Reports* **8**:1–11.

Uusi-Mäkelä, M.I.E. *et al.* (2018). Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio). *PLOS ONE* **13**:e0196238.

Venuti, A. *et al.* (2011). Papillomavirus E5: The smallest oncoprotein with many functions. *Molecular Cancer* **10**:1–18.

Verkuijl, S.A.N. and Rots, M.G. (2019). The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. *Current Opinion in Biotechnology* **55**:68–73.

de Villiers, E.M. (2013). Cross-roads in the classification of papillomaviruses. *Virology* **445**:2–10.

Vishnoi, K. *et al.* (2016). Cross-talk between Human Papillomavirus Oncoproteins and Hedgehog Signaling Synergistically Promotes Stemness in Cervical Cancer Cells. *Scientific Reports* **6**:1–15.

Wallace, N.A. *et al.* (2017). High-Risk High-Risk Alphapapillomavirus Oncogenes Impair the Homologous Recombination Pathway Banks, L. ed. *Journal of Virology* **91**:e01084-17.

Wang, H.K. *et al.* (2009). Robust production and passaging of infectious HPV in squamous epithelium of primary human keratinocytes. *Genes and Development* **23**:181–194.

Wang, J.W. and Roden, R.B.S. (2013). L2, the minor capsid protein of papillomavirus. *Virology* **445**:175–186.

Wang, T. *et al.* (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science (New York, N.Y.)* **343**:80–84.

West, A.M. *et al.* (2019). A conserved filamentous assembly underlies the structure of the meiotic chromosome axis. *eLife* **8**:1–27.

Whitehurst, A.W. (2014). Cause and Consequence of Cancer/Testis Antigen Activation in Cancer. *Annual Review of Pharmacology and Toxicology* **54**:251–272.

Williams, V.M. *et al.* (2011). HPV-DNA integration and carcinogenesis: Putative roles for inflammation and oxidative stress. *Future Virology* **6**:45–57.

Wojtasz, L. *et al.* (2009). Mouse HORMAD1 and HORMAD2, two conserved meiotic chromosomal proteins, are depleted from synapsed chromosome axes with the help of TRIP13 AAA-ATPase. *PLoS genetics* **5**:e1000702–e1000702.

Xiao, Q., Guo, D. and Chen, S. (2019). Application of CRISPR/Cas9-Based Gene Editing in HIV-1/AIDS Therapy. *Frontiers in Cellular and Infection Microbiology* **9**:69.

Xu, B. *et al.* (2013). Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas. *PLOS ONE* **8**:e66693.

Yamano, S., Dai, J. and Moursi, A.M. (2010). Comparison of Transfection Efficiency of Nonviral Gene Transfer Reagents. *Molecular Biotechnology* **46**:287–300.

Yang, F. *et al.* (2006). Mouse SYCP2 is required for synaptonemal complex assembly and chromosomal synapsis during male meiosis. *The Journal of Cell Biology* **173**:497 LP – 507.

Yim, E.-K. and Park, J.-S. (2005). The Role of HPV E6 and E7 Oncoproteins in HPV-associated Cervical Carcinogenesis. *Cancer Research and Treatment* **37**:319.

Yong, H.H. *et al.* (2009). The effect of MG132, a proteasome inhibitor on HeLa cells in relation to cell growth, reactive oxygen species and GSH. *Oncology Reports* **22**:215–221.

Yuan, L. *et al.* (2002). Female Germ Cell Aneuploidy and Embryo Death in Mice Lacking the Meiosis-Specific Protein SCP3. *Science* **296**:1115 LP – 1118.

Zheng, Z.-M. and Baker, C.C. (2006). Papillomavirus genome structure, expression, and post-transcriptional regulation. *Frontiers in Bioscience* **11**:2286–2302.