Identification of novel E2 binding sites within the HPV genome and their function in the regulation of viral gene expression

By

Mahamudul Hasan

A thesis submitted to the University of Birmingham for the degree of MRes in Cancer Sciences

UNIVERSITY OF BIRMINGHAM

School of Cancer Research
College of Medical and Dental Sciences
University of Birmingham
August 2014
Abstract

The human papillomavirus (HPV) E2 protein regulates the virus life cycle by modulating viral transcription and replication. To execute these functions, E2 binds to consensus sequences within the long control region (LCR). The possibility of novel E2 binding sites outside the LCR is the primary focus of this study. Moreover, the cellular protein CCCTC binding factor (CTCF), is known to regulate viral gene expression, therefore viral genome was screened for the presence of CTCF binding sites to see if they overlap the E2 binding sites. A comparison of CTCF expression within patient tonsil sections comprising normal (non-cancerous), HPV positive and HPV negative cancers may provide valuable information on the viral life cycle as well as disease progression.

Chromatin Immunoprecipitation assays using primary human tonsil keratinocytes containing episomal HPV16 genomes revealed novel E2 binding sites within the viral genome. Peak binding at base pairs 4400, 4500, 5600 and 6000 was detected, however consensus E2 binding sites do not exist in this region. CTCF was observed to bind to the same regions as HPV16 E2. Further investigation revealed a physical association between E2 and CTCF suggesting that CTCF could recruit E2 to the late region of the HPV genome. In patient tumour samples high levels of CTCF expression were observed throughout the epithelium; In contrast, the pattern of CTCF expression in the normal tonsil epithelium showed high expression in the lower layers that was dramatically reduced in the differentiated layers. Interestingly increased CTCF expression was observed in all areas of HPV positive tonsil sections in comparison to HPV negative sections. This difference in CTCF expression may be associated with HPV infection and important for the viral life cycle.
ACKNOWLEDGEMENTS

I would like to thank Dr Jo Parish for offering this wonderful project. Her constant support during my project as well as during the write up has been astronomical. I would like to thank her for having confidence in my ability. I would also like to thank Dr Katherine Feeney, Dr Karen Campos-León, Leanne Harris, Abida Siddiqa and Leisha Pentland for their constant support during my time in the lab as well as the write up period. I would also like to thank Dr Sally Roberts for the primary human tonsil keratinocytes. I would like to thank members of the InHANSE team especially Dr Davy Rapozo and Dr Sandra Ventorin von Zeildler for introducing me to the immunohistochemistry technique. I would also like to thank Dr Malgorzata Wiench and Harmeet Gill for helping with the ChIP-Seq library preparation.
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>2</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>3</td>
</tr>
<tr>
<td>CONTENTS</td>
<td>4</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>6</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>8</td>
</tr>
<tr>
<td>Overview of papillomaviruses</td>
<td>8</td>
</tr>
<tr>
<td>Epidemiology</td>
<td>10</td>
</tr>
<tr>
<td>Prophylactic action against papillomavirus</td>
<td>12</td>
</tr>
<tr>
<td>Vaccination against HPV</td>
<td>12</td>
</tr>
<tr>
<td>HPV genome</td>
<td>13</td>
</tr>
<tr>
<td>HPV life cycle</td>
<td>14</td>
</tr>
<tr>
<td>Viral oncoproteins E6 and E7</td>
<td>16</td>
</tr>
<tr>
<td>Early protein E5</td>
<td>17</td>
</tr>
<tr>
<td>Viral protein E2</td>
<td>17</td>
</tr>
<tr>
<td>Structure of E2</td>
<td>17</td>
</tr>
<tr>
<td>Function of E2</td>
<td>19</td>
</tr>
<tr>
<td>Viral protein E1</td>
<td>19</td>
</tr>
<tr>
<td>Viral protein E4</td>
<td>20</td>
</tr>
<tr>
<td>Late proteins L1 and L2</td>
<td>20</td>
</tr>
<tr>
<td>Control of HPV gene expression by the host cell protein CTCF</td>
<td>21</td>
</tr>
<tr>
<td>Structure of CTCF</td>
<td>22</td>
</tr>
<tr>
<td>HYPOTHESES AND AIMS</td>
<td>26</td>
</tr>
</tbody>
</table>
ABBREVIATIONS

- Adenocarcinoma (AC)
- Adenosine triphosphate (ATP)
- Antibody (ab)
- Base pair (bp)
- Binding sites (bs)
- Bovine papillomavirus (BPV)
- CCCTC binding factor (CTCF)
- Codons: Adenine, Uracil, Guanine (AUG)
- Cyclin depended kinases (CDK)
- Deoxyribonucleic acid (DNA)
- Early (E)
- Early polyadenylation (pAE)
- Epidermal growth factor receptor (EGFR)
- GlaxoSmithKline (GSK)
- Human papillomavirus (HPV)
- Late (L)
- Late polyadenylation (pAL)
- Long control region (LCR)
- Messenger RNA (mRNA)
- Mitogen-activated protein (MAP)
- Open reading frames (ORF)
- Origin of replication (Ori)
- Papillomavirus (PV)
- Polyadenylation (pA)
- Polypeptide protein tag (FLAG)
- Retinoblastoma protein (pRb)
- Squamous cell carcinoma (SCC)
• Zona occluden 1 (ZO-1)
INTRODUCTION

Overview of papillomaviruses

Small non-enveloped DNA tumor viruses such as papillomaviruses (PVs) infect a variety of vertebrate species \(^1\). PVs are ~55 nm in diameter \(^2\) and possess strict tissue tropism as well as host specificity, therefore human papillomaviruses (HPVs) solely infect squamous epithelium in the human anatomy \(^3\). More than 150 types of HPVs have been sequenced so far and, on the basis of DNA sequence analysis, they are categorised into \(\alpha\), \(\beta\), \(\gamma\), \(\mu\) and \(\nu\) PVs \(^4,5\) (Fig 1). The vast majority of the adult population are exposed to HPVs, which can infect an array of epithelial surfaces ranging from oropharyngeal to anogenital regions \(^6\) (Fig 2) \(^7\). The subclasses of PVs mentioned above possess different characteristics and pathological association \(^8\). \(\beta\) and \(\gamma\) PVs genera only seems to cause asymptomatic infections within immunocompetent individuals \(^9\). All mucosal HPV types belong to the \(\alpha\) PV genus \(^10\) and this subclass of HPV are well studied since they have been shown to be the etiological agents of cervical cancer \(^10\). Upon entry, the virus can cause cutaneous or mucosal skin infection, which may resolve spontaneously or develop into either benign or malignant tumors \(^10\). On the basis of their tumourigenicity, \(\alpha\) PVs can be divided into low-risk types such as HPV 6 and 11 and high-risk types such as HPV 16 and 18, low-risk HPV types are known to be the causative agent of benign warts and they are also capable of causing cutaneous lesions, however they are not known to cause neoplasia \(^11\). Conversely high-risk HPV are associated with cervical cancer \(^1,12,13\).
Figure 1 illustrates the phylogenetic tree of HPV. HPVs can be divided into five evolutionary groups with various epithelial tropisms and disease association. αPVs are comprised of low-risk mucosal types (orange shaded segment) known to cause benign warts and high-risk mucosal types associated with cervical neoplasia and cancer (pink shaded segment). Cutaneous HPVs are not known to be associated with cancer, α type (grey segment), β type (green segment) and γ type (blue segment). Image taken from 4.
**Epidemiology**

Cancer is one of the major causes of mortality and there are certain viral infections that are capable of predisposing infected individuals towards this pathological condition. Understanding how infectious agents are able to initiate and develop aberrant cellular proliferation opens up the potential to prevent or treat cancer with greater efficiency. Cervical carcinoma is the second most leading cause of cancer-related death amongst women worldwide. HPVs are accountable for 1.6% of all the cancer cases within UK. High-risk HPV types 16 and 18 have been identified as the predisposing factors in 70% of the cases of cervical cancer. It takes several years for a primary viral infection to develop into cervical cancer. Upon infection with HPV, the disease follows a progression route from low-grade to high-grade cervical lesions prior to carcinogenic transformation. Although cervical cancer is the most common cancer caused by HPV infection, head and neck and anogenital cancer cases are also known to be caused by HPV infection (Fig 2).

![Figure 2](image-url)

Figure 2 represents the range of cancers associated with HPV. Image taken from 7
The mechanism through which infections with high-risk HPV types persist for many years is still unknown. Cervical carcinoma can be divided into adenocarcinoma (AC) and squamous cell carcinoma (SCC). Around 85% of cervical cancer is SCC and 15% is AC. The prevalence of cervical cancer is higher in developing countries accounting for 15% out of all types of cancers. Increased rates of ACs are now being observed in developing countries. This could be due to the routine of screening programs that are capable of detecting AC. Current estimation predicts HPV prevalence amongst women without cervical alterations is 11.7% worldwide. High incidences of cervical cancer are observed in developing regions of the world for example in Kenya, Zimbabwe and Mozambique at 33.6%. In the Caribbean regions such as Trinidad and Tobago 35.4% and in Latin American regions 16.1% prevalence are observed. In contrast lower prevalence of HPV is observed in well-developed regions. For instance, western Europe, north America, UK and Asia show only 9%, 4.7%, 10% and 9.4% respectively.

Figure 3 is a representation of HPV prevalence worldwide. Red coloured regions correspond to high HPV prevalence, yellow coloured region correspond to intermediate HPV prevalence and green coloured region correspond to low HPV prevalence.
**Prophylactic action against papillomavirus**

Cervical screening programmes have significantly reduced the incidence of invasive cervical cancer and the associated mortality rate. However, the impact of cervical screening programmes in developing countries is not proving to be as effective. This could be due to a number of reasons such as inadequate population coverage or substandard cytology techniques. Women that are over the age of 30 are at greatest risk of developing cervical cancer due to persistent HPV infections. Therefore, the introduction of a cervical screening program could prove to be advantageous and aid towards early detection of cervical dysplasia in developing countries. A high incidence of HPV infection is observed amongst women under the age of 30, however these women do not exhibit any symptoms for a long period of time. Since the screening program fails to detect the presence of HPV, a prophylactic vaccine for women under the age of 20 was developed and could prove to be effective in preventing malignant transformation.

**Vaccination against HPV**

Vaccines against HPV were the first generated anti-cancer vaccine. Currently there are two prophylactic vaccines available against certain variants of HPV. Cervarix (GlaxoSmithKline) is a bivalent vaccine capable of protecting against high-risk HPV types 16 and 18 whereas a quadrivalent vaccine Gardasil (Merck) targets HPV types 6, 11, 16 and 18. These vaccines initiate the host immune response mediated via antibodies against the viral protein L1. Administration of Cervarix has shown to provide antibody responses up to 6.4 years on the other hand the quadrivalent vaccine has been shown to provide effective antibody responses for 5 years. However, additional preventive measures such as a screening program, lifestyle changes and physical examination should not be abandoned in vaccinated populations as results from a phase 3 trial reveal the inadequacy of the vaccines.
towards existing HPV infection and its progression towards pathological conditions
29.

**HPV genome**

The viral genome is 8000 basepairs (bp) in size with eight open reading frames (ORF), divided into three distinct territories termed the long control region (LCR), the early region and the late region 30. The LCR is approximately 850 bp long, that has binding sites for viral proteins E2 and E1 and cellular transcription factors as well as the origin of replication (ori) (Fig 4) 2,12,13. The early region of the genome comprises of six ORF which encode the viral proteins E1, E2, E4, E5, E6 and E7 2,30. Proteins encoded from the early region are mainly involved in maintenance of the genome, cellular growth promotion and replication of viral DNA 22. The structural proteins L1 and L2 are encoded by the late region of the viral genome and these proteins are the viral capsid proteins important in the assembly of infectious viral particles 31.
Figure 4 is an illustration of the HPV16 genome. The papillomavirus genome is a circular double-stranded molecule of 8,000 base pairs. There are three distinctive regions within the genome they are the LCR, the early (E) gene region, and the late (L) gene region. The genome encodes essential proteins for various stages of the viral life cycle. Early genes that are coloured in red represent oncoproteins and other early proteins are coloured in green. Late genes are coloured in orange. Early polyadenylation site (PAE) is located upstream of E5 protein and late polyadenylation site (PAL) located upstream of L1. Image taken from 4

**HPV life cycle**

A typical PV infection starts in the basal epithelial layer 32 where the virus gains access through a small wound or abrasion 8,31 (Fig 5). It has been shown that PV bind to heparan sulphate proteoglycans (HSPG) on the basal epithelial layer and the infection phase commences 32. The viral genome is maintained as a low copy number episome within infected basal cells, which act as a reservoir of infection 4. Initially viral genomes are replicated with the aid of host DNA replicative machinery of the basal epithelial layer 33. Expression of low viral proteins in undifferentiated cells enables HPV to evade the host immune system and maintain its infectious state 31. In the basal layer the episomal viral genome is maintained through the expression of E1, E2, E6 and E7 viral proteins. These proteins are crucial for the initiation of PV
life cycle since, replication of viral DNA is facilitated by the E1 and E2 proteins. E6 and E7 regulate replication competence either through the degradation of tumor suppressor protein p53 by E6 or through the inactivation of retinoblastoma protein (pRb) by E7. Expression of E6 and E7 are down regulated by viral transcription factor E2. Basal differentiation causes the activation of late viral functions, which are mediated by E4, E5 and E1. Viral DNA replication and protein expression are amplified upon late promoter activation, this leads to an increase in viral copy number.

Figure 5 illustrates the viral life cycle in cervical epithelium. The HPV life cycle initiates following infection of the epithelial cells in the basal layer. Cells with red nuclei represent virally infected basal epithelial cells that are dividing. Viral protein E4 expression is up regulated in cells residing in the middle layers, this protein is essential for the amplification of the genome. Green coloured cells are E4 positive with red nuclei. Genome amplification in the middle layers is aided by E6 and E7 that inhibit tumor suppressor proteins p53 and pRb, respectively. In the upper layers virus particles are packaged and released from cornified epithelium.

Figure 5 illustrates the viral life cycle in cervical epithelium. The HPV life cycle initiates following infection of the epithelial cells in the basal layer. Cells with red nuclei represent virally infected basal epithelial cells that are dividing. Viral protein E4 expression is up regulated in cells residing in the middle layers, this protein is essential for the amplification of the genome. Green coloured cells are E4 positive with red nuclei. Genome amplification in the middle layers is aided by E6 and E7 that inhibit tumor suppressor proteins p53 and pRb, respectively. In the upper layers virus particles are packaged and released from cornified epithelium.

Persistent infection drives cell proliferation in the basal and the parabasal cell layers. Infected cells within the basal cell layer undergo mitosis where daughter cells are produced, then migrate towards the epithelial surface and begin to differentiate. In the upper epithelium expression of E6 and E7 enables the HPV infected cell to re-enter S-phase of the cell cycle, leading to an increase in copy number of the viral
Down regulation of these oncoproteins causes the release of tumor suppressor protein pRb and p53 that was previously inhibited by E6 and E7. This enables host cells to progress through normal differentiation followed by the expression of late viral proteins L2 and L1. The viral structure is formed by capsomers containing the minor capsid protein L2 and major capsid protein L1. Viral genome becomes encapsulated in the nucleus which is followed by the release of complete virions.

**Viral oncoproteins E6 and E7**

The oncoproteins E6 and E7 are the main causative agents in the development of HPV induced cervical cancer. E6 consists of two zinc finger motifs and remains localised in the cytoplasm as well as the nucleus of the infected keratinocytes. Interference of E6 in cell cycle regulation through p53 ablation is well documented. The G1/S and G2/M phases of the cell cycle are regulated by p53 and this tumor suppressor protein becomes activated in response to cellular stress and DNA damage response. p53 degradation is initiated by E6 and the E6-associated protein (E6-AP) which target p53 for proteasome-dependent degradation by conjugation of ubiquitin. Degradation of p53 eliminates restraints on DNA synthesis enabling replication of viral DNA. In addition, E6 has the capacity of upregulate and sustain telomerase activity over numerous cell divisions and it has recently been established that E6 degrades PDZ substrates and contributes to tight junction ablation via zona occludens 1 (ZO-1) relocation. PDZ motifs are associated with cell signalling, polarity and proliferation.

E7 consists of three conserved regions and the conserved region 2 contains a LXCXE motif. This motif mediates the binding of E7 to the tumor suppressor protein pRb. pRb is hypophosphorylated thus enabling the inhibition of the transcription factor E2F during the G1 phase of the cell cycle. On the other hand, hyperphosphorylation of pRb by cyclin dependent kinases (CDK) induces the
dissociation of E2F from pRb and drives cell cycle progression \(^{60}\). Association of E7 to pRb causes its degradation, therefore, the inhibitory effect on E2F imposed by pRb is alleviated and early S-phase entry is promoted \(^{35,61}\).

**Early protein E5**

E5 is a hydrophobic protein that localises to the endoplasmic reticulum, Golgi apparatus and plasma membrane \(^{62,63}\). Within a cell culture system HPV E5 exhibits weak transformation ability \(^{64}\). It has also been found that the functions of E6 and E7 are reinforced by E5 and there is cumulative evidence on E5 increasing the half-life of epidermal growth factor receptor (EGFR) \(^{65,66}\). The interaction between EGFR and HPV16 E5 results in an increased activation of MAP kinases \(^{67,68}\).

**Viral protein E2**

E2 is DNA binding protein encoded by all PVs in the early and intermediate stages of the viral life cycle. E2 regulates HPV genome transcription and viral DNA replication \(^{69}\). DNA binding protein E2 is sequence specific and capable of binding to 12 base pair motifs that are located in the LCR \(^{70}\). The E2 protein is comprised of an N-terminal transactivation domain linked to a DNA binding/dimerization C-terminal domain \(^{71}\) via a flexible sequence known as hinge \(^{72}\). The full length E2 protein forms dimers that are capable of initiating replication and facilitate transcription \(^{73,74}\).

**Structure of E2**

The N-terminal region or transactivation domain of E2 forms a characteristic “cashew shaped structure” \(^{72}\) across all PVs. Structural analysis of HPV 16 E2 reveals three long \(\alpha\)-helices are arranged anti-parallel to each other within the transactivation domain, which are thought to form a protein interaction surface. On the other hand, anti-parallel \(\beta\)-sheets that form an unusual \(\beta\)-barrel structure
surrounded by four α-helices make up the C-terminal DNA binding and dimerization domain. The hinge region is thought to be relatively unstructured and serves as a flexible linker between the N- and the C-terminal domains of the protein (Fig 6).

Figure 6 illustrates HPV 16 E2 with transactivation domain with α-helices arranged anti-parallel with each other and DNA binding domain with anti-parallel β-sheets are observed. The two domains are connected via hinge structure.

The transactivation domain contains essential residues that orchestrate transcriptional regulation and these residues are located on the outer surface of the domain. In contrast, residues that are associated with replication are located on the inner surface of the domain, because they disrupt binding to E1. Unlike the transactivation domain sequence specificity does not perturb the DNA binding domain nevertheless it is known to stabilise interactions between DNA and protein.
**Function of E2**

E2 is a sequence specific DNA binding protein which recruits cellular factors to the viral genome which are capable of either activating or repressing transcriptional processes \(^78\). E2 binds to DNA in a sequence specific manner. There are four well characterised E2 binding sites within the LCR of the \(\alpha\) PV genome \(^10\). The binding of E2 to DNA is dependent on the consensus binding sequences they are ACCG(N)\(_4\)CGGT or ACC(N)\(_6\)GGT where N represents a spacer region often rich in T or A \(^1,69\). Transcriptional activation or repression of viral gene expression is dependent on the E2 binding sites and recruitment of E2-associated cellular factors \(^79,80\). There is evidence of transcriptional repression via competitive binding of short forms of E2 to the E2 binding sites \(^72,81\). The short forms of E2 are thought to dimerize with full length E2 to further enhance transcriptional repression \(^82\). Viral DNA replication is initiated when E1 is loaded on to the replication origin via association with E2 \(^83,84\). Therefore, E2 plays a supporting role in the replication mechanism mediated via E1 \(^83,85\). Replication of viral DNA takes place in nuclear foci and the constructions of these foci are dependent upon E2 protein \(^86,87\), although the exact make-up and nature of these foci is at present unclear. In order to facilitate viral genome maintenance, retention and partitioning E2 tethers viral genomes to the host chromosome by association with chromatin bound cellular proteins \(^88,89\). However in comparison E2 proteins of the \(\alpha\) PV species exhibits a weak binding to host chromosomes \(^90\).

**Viral protein E1**

One of the most conserved proteins encoded by all PV is the E1 protein, which is an adenosine triphosphate (ATP)-dependent DNA helicase \(^91\). Conservation of the E1 protein is indicative of its role as a replicative helicase, which delineates conscientious replication of the viral episome \(^92\). E1 activity is essential for viral
genome replication, copy number increase and regulation of episomal levels. At the viral ori, E1 constructs into a double-hexamer that melts DNA at the origin prior to establishment of the replication fork.

**Viral protein E4**

E4 ORF encodes a protein that is variable in size between PV types. Expression of E4 protein arises from the spliced mRNA product $E1^{\wedge}E4$ that consist of the E1 initiation codon and few sequences from the E1 ORF. E4 is encoded from the early region of the HPV genome however it is known to contribute predominantly to the late stages of the viral life cycle, for instance amplification of the genome and capsid protein expression, and expression is only detected in differentiated epithelium. E4 has been shown to contribute to keratin network disruption and which potentially aids towards viral transmission.

**Late proteins L1 and L2**

L1 and L2 are viral capsid proteins transcribed from the late region of the viral genome. L1 and L2 are expressed towards the end of the viral life cycle. L2 is the minor capsid protein and recent studies have established its role in HPV genome encapsidation. L2 plays an essential role during viral entry into host cells, for instance L2 disrupts subcellular trafficking, endosomal membranes and initiates conformational changes on the virions that are attached to the cell. L1 is the major viral capsid protein. Capsid proteins are synthesised initially within the cytoplasm after which they migrate to the nucleus where viral chromatin is packaged. L1 capsomeres are packaged in the cytoplasm.

Upon re-infection, the viral L1 protein interacts with HSPG that is expressed on the surface of basal cells of the epithelium. This interaction causes a change in conformation of the major capsid protein and leads to the exposure of minor capsid protein. The minor capsid protein is then subjected to furin cleavage, which
allows the virus to attach itself to an alternative surface receptor and infect the cell

The functions of E2 in the control of viral gene expression and replication by its
association with consensus binding sites within the LCR have been well
documented. However, additional E2 binding sites outside of the viral LCR exist. The
function of these binding sites and whether E2 associates with them in vivo has
not been studied and requires further analysis.

**Control of HPV gene expression by the host cell protein CTCF**

Several cellular proteins have been shown to be recruited to the HPV genome to
control early and late gene expression. CCCTC-binding factor (CTCF) is an 11
zinc-finger protein that has been shown to regulate HPV gene expression by
recruitment to consensus binding sites within the HPV18 genome (Parish,
unpublished). CTCF is normally located in the linker regions of human chromatin that
are circumscribed by nucleosomes and it was initially identified as transcriptional
repressor for chicken c-myc and lysozyme genes. CTCF has also been shown
to function as a transcriptional activator. When CTCF is positioned between a
gene promoter and an enhancer it has the capability to harbour insulator activity.
This halts communication between the promoter and the enhancer and therefore
blocks transcriptional activation. CTCF is associated with a variety of molecular
functions depending on the genetic locus involved, such as transcriptional activation,
transcriptional repression and enhancer blocking activity. CTCF is
involved in a variety of long range processes such as chromatin looping,
chromatin insulation, chromosome segregation, and nuclear organisation.
CTCF binding sites are highly conserved across different cell types. Interspecies
CTCF binding profile within the liver was investigated, which unveiled 5000 highly
conserved sites between species and tissues.
**Structure of CTCF**

The complete structure of CTCF comprises of 3 major functional parts an N-terminal region, a DNA binding middle region and the C-terminal region. CTCF protein is subjected to post-translational modification such as poly-(ADP)-ribosylation (PARylation) at the C-terminus and phosphorylation at the N-terminus. CTCF is also SUMOylated. Phosphorylation at the N-terminal region switches CTCF function from transcriptional repressor to transcriptional activator. PARylation is crucial for CTCF to function as an insulator and the lack of this modification abolishes barrier activity. SUMO Small ubiquitin like protein is capable of modifying CTCF which enhances the CTCF repressor function at the c-myc P2 promoter.

CTCF binds to DNA via a poorly defined consensus binding motif. A single CTCF binding site consists of a primary and secondary motif. These binding motifs are recognised by CTCF via its 11 zinc fingers. Within certain locations the primary binding motif is capable of defining a CTCF binding site whereas, the secondary binding motif enhances CTCF binding. Bases that are adjacent to primary and secondary motifs are also thought to be important for CTCF binding.
Figure 7 represents CTCF structural features as well as substitution of amino-acid specific to tumours within zinc fingers. CTCF protein present in humans has a DNA-binding domain, which comprises of ten $C_2H_2$-class ZFs (ZFs 1–10) and one $C_2HC$-class ZF (C-terminal ZF11)\(^{160}\).

The 11 zinc fingers of the CTCF DNA binding region do not all commit equally to the binding of DNA\(^{140}\). Zinc fingers 4-7 target the core DNA binding motif and bind to 80% of the CTCF binding sites, whereas zinc fingers 8-11 and 1-2 stabilise CTCF broadly\(^{123}\). CTCF is capable of affecting DNA methylation status through the formation of DNA (cytosine-5)-methyltransferase 1 (DNMT1) and poly(ADP-ribose) polymerase 1 (PARP1) complex. Activation of PARP1 via CTCF, inactivates DNMT1 through poly(ADP-ribosyl)ation and therefore perpetuates methyl free CpGs in the DNA\(^{124,125}\). It has also been shown that binding of CTCF is inhibited through methylation of CTCF binding sites\(^{141,142}\) (Fig 8).
Figure 8 illustrates maintenance of CTCF and DNA binding with each other. Binding of CTCF binding is partially maintained via occupancy of the nucleosomes along with differential DNA methylation at specific CTCF binding sites. This suggests that the cells are able to remodel chromatin complexes by using ATP and consequently maintain certain CTCF binding sites along with the nucleosome occupancy at these sites. High levels of 5-methylcytosine are observed at CTCF binding sites this corresponds to a low CTCF occupancy. Methylation in this figure is depicted by filled red circles whereas open circles depicts unmethylated DNA.

CTCF has a major impact on the three dimensional structure of DNA which in turn has a major impact on gene regulation. Interaction of CTCF with nucleophosmin/B23 as well as the nuclear matrix, could be the factors contributing towards three dimensional DNA conformation. Partitioning and positioning of DNA inside the nucleus is also thought to be mediated by CTCF (Fig 9). CTCF is a unique protein and influences cellular processes such as insulation, transcription, gene activation and tumour suppression. All of these roles mentioned could be crucial in the maintenance of the HPV genome and HPV gene expression. Additionally, there is evidence that CTCF and cohesin interact which has been
shown to be essential for CTCF function \(^{146}\). E2 and cohesin have been shown to interact with one another within HPV (Parish unpublished).

Figure 9 illustrates the positioning and partitioning of DNA mediated by CTCF. Individual interphase chromosomes are formed from Double stranded DNA that occupies a certain area of the nuclear volume, this leads towards the formation of chromosome territories. Within the transcriptionally active centre of the nucleus CTCF constructs a non-random interchromosomal connections of certain loci. CTCF is known to bind at the borders of transcriptionally silent lamina associated domains (LADs) \(^{147}\).
HYPOTHESES AND AIMS

Hypotheses

• Alternative E2 binding sites exist in HPV genomes outside of the viral long control region (LCR) that are important in the regulation of viral gene expression

• Novel E2 binding sites co-localise with CTCF binding sites in the HPV genome and binding of E2 to these novel binding sites enhances CTCF-dependent regulation of viral gene expression

• Evaluation of HPV16 E2 and CTCF binding within the host genome could provide valuable information on viral genome maintenance and HPV life cycle.

• Differentiation dependent CTCF expression is important in the control of early and late gene expression in the virus life cycle. Alterations in the pattern of CTCF expression contributes to cancer-associated deregulation of viral gene expression.

Aims

• Identification of novel HPV 16 E2 binding sites via CHIP assay.

• Identification of CTCF biding sites co-localisation with HPV16 E2 binding sites.

• Determining the host binding sites for CTCF and HPV 16 E2, in order to deduce their role in viral genome maintenance and life cycle.
Analysis of differentiation dependent expression of CTCF in normal, HPV negative and HPV positive keratinocytes.
MATERIALS AND METHODS

Chromatin Immunoprecipitation (ChIP)

Active Motif ChIP-IT Express Enzymatic Kit catalogue number 5009 and 53035 was used to carry out the chromatin Immunoprecipitation experiment.

Chromatin preparation

Primary human tonsil keratinocytes with episomal HPV16 DNA was kindly provided by Dr Sally Roberts. Dr Jo Parish cultured the tonsil keratinocytes in 15 cm tissue culture dishes to 70-80% confluency. The tonsil keratinocytes were then fixed in 1% formaldehyde solution diluted in growth medium for 3 minutes at RT with gentle rocking. The cells were then washed in ice cold phosphate buffer saline (PBS) for 5 seconds. The formaldehyde was then quenched with using Glycine Stop-Fix solution and incubated for 5 minutes at RT. The cells were then washed with ice cold PBS and harvested by scraping with a rubber policeman in 1ml ice cold PBS. The cells were centrifuged at 600 x g for 10 minutes at 4°C. The supernatant discarded and the cell pellets stored at- 80°C.

Enzymatic Shearing of Chromatin

The pellet was resuspended in ice cold lysis buffer, supplemented with protease inhibitor cocktail (PIC) and phenylmethylsulfonyl fluoride (PMSF), and incubated on ice for 30 minutes. The resuspended cell pellet was transferred to the dounce homogeniser (Kimble-Kontes part number 885302-002 with tight fitting B pestle) and dounced on ice forty times. A sample was taken from the dounce homogeniser and the cells were observed under light microscope to ensure efficient disruption of the plasma membrane. The nuclei were pelleted at 2600 x g for 10 minutes at 4°C. The supernatant was discarded and the pellet resuspended in digestion buffer.
supplemented with PIC and PMSF and incubated at 37°C for two minutes. The enzyme shearing cocktail was diluted with 50% glycerol to give a 1:100 dilution and 2.5 ml of the diluted enzyme cocktail was added to each sample. Samples were incubated at 37°C for 10 minutes and digestion was then halted with 0.5M EDTA and the samples incubated on ice for 10 minutes. Subsequently the chromatin was centrifuged at 20000 x g at 4°C for 10 minutes, and the supernatant containing the sheared chromatin was collected A sample of the sheared chromatin was run on an agarose gel.

**Chromatin shearing by sonication**

Chromatin pellet was resuspended thoroughly in 600 µl SDS lysis buffer (1% SDS, 50mM Tris-HCl, pH 8.1 and 10mM EDTA) along with protease inhibitors, the reaction was incubated on ice for 30 minutes. Cell lysates were loaded in polystyrene conical tubes and sonicated in the Bioruptor (sonication bath) for 7 cycles of 30 seconds on and 30 seconds off. Sonicated chromatin was centrifuged at 18000 x g at 4°C for 10 minutes. Sonicated cell supernatant was diluted with ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris-HCl pH 8.1, 167mM NaCl and protease inhibitor) to reduce the final concentration of SDS to less than 0.1%. A sample of the sonicated chromatin was run on an agarose gel to check shearing efficiency.

**Immunoprecipitation of Chromatin**

The immunoprecipitation was constructed in a siliconised microcentrifuge tube. 25 µl protein G magnetic beads (Active Motif) was added along with 10 µl ChIP buffer 1 and ~7 µg of sheared chromatin, protease inhibitor cocktail, respective antibody and RNase free H₂O was added to make up a total reaction of 100 µl. This reaction mixture was left
rotate on an end-to-end rotator overnight at 4°C. The following morning the reaction mixture was briefly centrifuged and placed on the magnetic stand to pellet the beads at the side of the tube and the supernatant was discarded. The beads were subsequently washed once in 500 µl in ChIP buffer 1 and three times in ChIP buffer 2.

<table>
<thead>
<tr>
<th>Antibody Name</th>
<th>Antibody type</th>
<th>Isotype</th>
<th>Manufacturer</th>
<th>Catalogue number</th>
<th>Species</th>
<th>Volume µl</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV16 E2</td>
<td>Polyclonal</td>
<td>-</td>
<td>Parish Lab</td>
<td>-</td>
<td>Sheep</td>
<td>3</td>
</tr>
<tr>
<td>TVG 261</td>
<td>Monoclonal</td>
<td>IgG1</td>
<td>ABCAM</td>
<td>17185</td>
<td>Mouse</td>
<td>3</td>
</tr>
<tr>
<td>CTCF</td>
<td>Polyclonal</td>
<td>IgG</td>
<td>Active Motif</td>
<td>61311/61312</td>
<td>Rabbit</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1 outlines the primary antibodies used for ChIP

**Chromatin elution**

The washed beads were resuspended in 50 µl elution buffer AM2 and incubated for 15 minutes at room temperature on an end-to-end rotator. The beads were centrifuged briefly and 50 µl reverse cross-linking buffer added, the beads were then immediately placed on the magnetic stand and the supernatant containing the chromatin was removed and stored in a fresh Eppendorf tube. 5M NaCl and ChIP buffer 1 were added to 10% input DNA to a total volume of 100 µl and these samples along with the chromatin immunoprecipitation reactions were incubated at 95 °C for 15 minutes to reverse the crosslinks. Proteinase K solution was then added and incubated at 37°C for one hour. The action of Proteinase K was halted by Proteinase K stop solution. CHIP
DNA was cleaned using the Sigma-Aldrich PCR clean up kit catalogue number NA1020.

Real-time PCR

Real-time PCR reactions were set up in a 96 well plate. A master mix was prepared in a PCR hood. PCR primers were diluted to a final concentration of 10 pmol\( \mu l^{-1} \) of which 0.125 ml of forward and 0.125 ml reverse primer were added to each reaction along with 8.75 ml of H2O and 10 ml of Senimix SYBR No-ROX 2X (Bioline). The final concentration of the primers was 0.42 pmol\( \mu l^{-1} \). Input DNA was serially diluted and utilised to generate a standard curve. The ChIP DNA samples were plated in duplicates and amplified for 40 cycles. The thermal profile of the qPCR was set up an annealing temperature of 50°C for 45 seconds and an extension temperature of
72°C for 45 seconds. The primers used for qPCR reactions are listed in the appendix table 4. Real-time PCR was run using Stratagene MX3005P.

**Library preparation for ChIP-SEQ**

The library was prepared using NEBNext ChIP-SEQ library preparation Prep Master Mix Set for Illumina (# E7350) as per manufacturer's instructions. Index primer 4, 6 and 12 from the NEBNext Multiplex Oligos for Illumina was utilised to generate the library for ChIP-SEQ. (Appendix table no 5 index primer sequence)

**Cell culture**

C33a cells are an HPV negative human cervical cancer cell line. C33a cell lines were cultured in Dulbecco Modified Eagle Medium (DMEM) (Sigma-Aldrich) containing 10% foetal bovine serum (FBS) (Life Technologies). Cells were incubated at 37°C, 5% CO₂. Cell culture was performed in a laminar flow tissue culture hood.

**C33a cell line transfection**

Transfection of cells was performed using X-tremeGENE Transfection Reagent (Roche). Ten centimeter dishes were seeded with 2 x 10⁶ C33a cells in 10 ml of growth medium. Cells were transfected 24 hours later. 500 µl of serum free DMEM was added to 3 µg of HPV 16 E2 expressing plasmid pJ4Omega-16E2 which encodes full length HPV16 E2 under the control of a CMV promoter. This was mixed thoroughly and 6 µl of X-tremeGENE transfection reagent was added. The mixture was left for 15 minutes at room temperature before adding to the cells dropwise. The cells were incubated for 24 hours at 37°C, 5% CO₂ before harvesting.
**Cell lysis**

The 10 cm dishes were washed twice with PBS and scraped in 1 ml of ice cold PBS. The cells were collected and centrifuged at 1000 x g at 4°C for 10 minutes. The PBS was removed carefully and cells were lysed in 300 µl ice cold lysis buffer (50mM Tris-HCl pH 7.4, 100mM NaCl, 20mM NaF, 10mM KH2PO4, 1% Triton x-100, 0.1mM DTT, 10% glycerol, 1% protease inhibitor cocktail). The samples were incubated on ice for 30 minutes and the cell suspension was then sonicated at 30% amplitude for 10 seconds. The cell suspension was centrifuged again at 10600 x g at 4°C for 20 minutes to remove the debris. The lysate was collected and a 10% input sample is taken for gel analysis which is mixed with 6X SDS gel loading dye and boiled for 10 minutes at 95°C.

**Co-Immunoprecipitation**

Protein G conjugated sepharose beads (Sigma) were washed with binding buffer three times (Tris-HCl pH 7.4, 100mM KCl, 0.1 mM EDTA, 0.20% IGEPAL CA-630, 0.10% BSA and 2.50% DTT). 200µl of binding buffer, 200µl of cell lysate, 10µl of protein G conjugated sepharose beads and 2 µl of respective antibody were mixed. Table 3 outlines the respective antibodies used. This reaction mixture was left to rotate on the wheel overnight at 4°C.
The samples were briefly centrifuged at 6000 x g and the supernatant discarded. The protein G beads were washed three times in 500µl of wash buffer (100mM Tris-HCl pH 7.4, 100mM NaCl, 0.50% IGEPAL CA-630 and 2mM DTT). After the final washing step the beads were suspended in 2X SDS gel loading buffer and the samples boiled for 10 minutes at 95°C.

**Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting**

A 10% SDS polyacrylamide gel was poured and loaded with the immunoprecipitation reactions along with 10% inputs. The size marker used during this experiment was Pageruler plus protein ladder (Thermo scientific). SDS gels were run at 100 volts for the first 20 minutes and 120 volts for 50 minutes in 1 x running buffer. Proteins were transferred to a PVDF membrane (Roche Diagnostic) in 1X transfer buffer (125mM Tris-HCl pH 8.3, 1.25M glycine and 5% methanol). The transfer was carried out at 100V and 400mA for 1 hour and 10 minutes. The transferred membrane was blocked overnight in 5% milk in Tris buffered saline (10mM Tris-HCl pH 7.6, 150mM NaCl), 0.05% tween 20 (TBS/T) to prevent non-specific binding of the detecting antibodies, at 4°C with gently rocking. 5% milk was prepared using TBS/T. The membrane was incubated in primary antibody diluted in

<table>
<thead>
<tr>
<th>Antibody Name</th>
<th>Antibody type</th>
<th>Isotype</th>
<th>Manufacturer</th>
<th>Catalogue number</th>
<th>Volume µl</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV16 E2</td>
<td>Polyclonal</td>
<td>Parish lab</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>Sheep</td>
</tr>
<tr>
<td>TVG 261</td>
<td>Monoclonal</td>
<td>IgG1</td>
<td>ABCAM</td>
<td>17185</td>
<td>2</td>
<td>Mouse</td>
</tr>
<tr>
<td>IgG</td>
<td>-</td>
<td>IgG</td>
<td>SANTA CRUZ</td>
<td>sc-2025</td>
<td>2</td>
<td>Mouse</td>
</tr>
<tr>
<td>CTCF</td>
<td>Polyclonal</td>
<td>IgG</td>
<td>Active Motif</td>
<td>61311/61312</td>
<td>2</td>
<td>Rabbit</td>
</tr>
</tbody>
</table>

Table 2 outlines the antibodies used to perform co-immunoprecipitation
5% milk buffer for one hour at room temperature by gently rocking. Table 4 lists the primary antibodies used.

![Table 4: Primary Antibodies Used](image)

Table 3 outlines the primary antibodies used during western blot detection.

Detection of the membrane

Subsequently the membranes were washed in 1X TBS/T for 5x five minutes and the membranes incubated with horseradish peroxidase (HRP) conjugated secondary antibody incubation diluted 1:10,000 in 5% milk buffer for one hour at room temperature. The membranes were then washed 5 times five minutes in 1X TBS/T. The membrane was then incubated with enhanced chemiluminescence western blotting substrate (Thermo scientific catalogue number 32106) for three minutes. The membrane was then immediately placed in a plastic sleeve and detected in the Fusion Fx7 (Vilber Lourmat) machine by using FusionCapt Advance software. The membrane was stripped for the purpose of reprobing by washing 4 x 10 minutes with 10% acetic acid and 5 x 5 minutes in TBS/T.

Immunohistochemistry

HPV positive and HPV negative tonsil tumour sections were selected. The slides were deparaffinised and rehydrated, this was performed to improve the staining quality. The slides were placed in Xylene for 4x five minute incubations. Subsequently the slides were incubated in 100% isopropanol for three minutes. This
step was repeated three times. This was followed by incubation in 90% isopropanol for a further three minutes. The slides were rinsed in sterile deionised water and kept wet.

Formalin fixed tissue requires an antigen retrieval step before immunohistochemistry staining can be proceeded. Heat-induced epitope retrieval is performed using EMS Retriever 2100 for 30 minutes. The retriever was filled with distilled water. The antigen retrieval solution was manufactured by DAKO (product number S1699) pH 6.0 and the 10x antigen retrieval solution was diluted in deionised water to 1x.

The slides were marked with a Novocastra NovoPen (hydrophobic pen) this marginalises the area to be stained. The slides were washed with 1X TBS wash buffer and immunohistochemistry staining was performed using a Novolink Polymer Detection System (Leica).

The slides were subsequently subjected to peroxidase block for five minutes, this neutralises endogenous peroxidase. The slides were washed twice with 1X TBS before CTCF antibody incubation overnight at 4°C. The antibody was diluted in 1X TBS as detailed in the results section. 500 µl of antibody was used to cover the entire tissue section.

After 24 hours, the slides were washed with 1X TBS three times before incubating with post primary solution for half an hour. This was followed with 1X TBS washing step before incubating with Novolink polymer for half an hour. The slides were washed again with 1X TBS before developing peroxidase activity with 3,3′-diaminobenzidine (DAB) working solution. DAB working solution was prepared in a 1:50 dilution factor and 500 µl of DAB working solution was used to cover the entire section for five minutes. The slides were washed with deionised water and counter stained with haematoxylin and rinsed with deionised water.
The slides were dehydrated in 90% isopropanol followed by 100% isopropanol and Xylene before being mounted.
RESULTS

Assessment of E2 association with the HPV 16 genome by ChIP

Human papillomavirus E2 is a sequence specific DNA binding protein which binds to specific consensus sequences (ACCG(N)₄CGGT or ACC(N)₆GGT) \(^{73}\). In high-risk HPV genomes, there are four highly conserved E2 binding sites located in the viral locus control region \(^{151}\). However there are alternative binding sites, which lie outside the viral locus control region and are crucial in the regulation of viral gene expression. In order to investigate this, both forward and reverse primers were designed, that cover the entire HPV 16 genome (see appendix table number 4). Primary human tonsil keratinocytes containing episomal HPV 16 genome was cultured and chromatin from the cultured cell lines was extracted. The extracted chromatin was either sheared via sonication or enzymatic cocktail (Fig 11).
Sheared chromatin was immunoprecipitated with FLAG antibody, sheep HPV 16 E2 antibody, mouse HPV16 E2 antibody TVG 261 or rabbit CTCF antibody and co-precipitation of the HPV16 genomes analysed using real-time PCR.

Before normalising the data obtained from real-time PCR the efficiency of the amplified DNA was checked against a standard curve, generated by a serially diluted sheared chromatin. Ten-fold serial dilution was prepared to generate the standard curve, which was run alongside the unknown ChIP DNA. Unknown DNA was compared to the known standards which allowed the evaluation of the quantitative data, (Fig 12)
This determines whether the primers used in the reactions are amplifying a specific region or is there non-specific amplification. If the primers are amplifying a specific region of DNA then this would be corresponded by a single peak and if the primers are amplifying non-specific regions then there would be more than one peaks (Fig 13). No template control (NTC) reactions were run to detect any possible contamination of reagents or primer dimer formation. Detection of DNA amplification within the NTC samples lead to no further analysis of the data set and the real-time PCR was repeated with freshly diluted primers.
Figure 12 represents a standard curve generated by serially diluted sheared chromatin. Unknown ChIP DNA is represented by ▲ and standards are represented by ▣. A RSq value of 0.992 was noted along with 99.6% efficiency. Initial quantity (copies) vs Ct(dR) was used to plot the standard curve.

Figure 13 represents a dissociation curve with products aligning as a single peak. Temperate (°C) verses fluorescence (-R¹ CT) value was used to plot the graph.
The experiment shown in (Fig 14) was performed with high passage of tonsil keratinocytes. Enzymatic shearing of the chromatin was utilised as mentioned earlier. Using FLAG, HPV 16 E2, TVG 261 and CTCF antibodies the sheared chromatin was immunoprecipitated and binding of E2 and CTCF was analysed by real-time PCR. Raw data obtained from real-time PCR analysis was normalised to FLAG. Figure 14 represents graphical analysis of the data obtained from the real-time PCR. There are two distinct peaks are observed one of which is between 4000 to 5000 bp and the other peak is around 6000 bp.

Binding of both CTCF and HPV 16 E2 are superimposed on one another. Nevertheless, in the late region of the genome, CTCF binding was higher compared to HPV 16 E2. Binding of commercially available HPV 16 E2 antibody TVG 261 was higher in the early region of the HPV genome, specifically in with the viral LCR. Binding in the late HPV genome was not observed experimentally.
Figure 14 Real-time PCR analysis of immunoprecipitated chromatin derived from primary human tonsil keratinocytes with episomal HPV 16 genomes. Chromatin was sheared by enzymatic digestion. Samples were immunoprecipitated with HPV 16 Sheep E2 antibody (Parish Lab), HPV16 E2 TVG 261 antibody and CTCF antibody. Primers covering the entire HPV 16 genome were utilised to assess binding of E2 and CTCF throughout the genome. Samples were analysed in duplicate and an average was calculated. Data for each immunoprecipitation was normalised to FLAG antibody (negative control). This experiment was repeated three times.

The sonication method was implemented in order to improve the efficiency of the chromatin preparation and shearing. Subsequent treatment of the chromatin remained the same for the rest of the experiment. Graphical analysis of the ChIP DNA revealed binding of HPV 16 E2 in the early region and at around 4400, 5600 and 7200 bp with the polyclonal sheep E2 antibody (Fig 15). Two distinct peaks were observed one of which was at 4400 base pairs and the other one at 5600 base pairs and binding of CTCF and HPV 16 E2 is superimposed one another. With the
sonicated ChIP DNA additional HPV 16 E2 and CTCF binding site was identified within the LCR.

**Primary human tonsil keratinocytes with episomal HPV16 genomes**

![Graph](image)

Figure 15 represents the real-time PCR analysis of the chromatin derived from primary human tonsil keratinocytes with episomal HPV 16 genomes. Chromatin was sheared with enzymatic cocktail. Samples were immunoprecipitated with HPV 16 Sheep E2 antibody (Parish Lab), TVG 261 antibody and CTCF antibody. Primers covering the entire HPV16 genome was utilised during the course of the experiment. Samples were loaded in duplicates and an average was calculated. Data set was normalised to FLAG antibody, this is a negative control antibody.

**End repaired ChIP DNA**

End repaired ChIP DNA was run on a bioanlyser in order to determine the fragment size of the library, as well as the concentration of the end repaired ChIP DNA required for ChIP-SEQ. Unfortunately the fragment size of my library was inconsistent in comparison to its original size. In addition the concentration of the end repaired ChIP DNA was very low, therefore the ChIP-SEQ analysis had to be
abandoned and there was insufficient time to repeat the ChIP analysis to improve the yield of the library.

Co-Immunoprecipitation of C33a cells transfected with HPV 16 E2

From the data analysis it is evident that both HPV 16 E2 and CTCF are co-localizing with one another on the HPV16 genome. However it is not clear from the real-time PCR analysis which protein is binding to the DNA first and co-localizing with the other. From previous work in the Parish laboratory, it is known that there is a cluster of CTCF binding sites within the late region of the HPV16 genome and consensus E2 binding sites do not exist in this region. Therefore, it was hypothesised that CTCF recruits E2 to the HPV genome via a protein-protein interaction. In order to further establish this interaction a co-immunoprecipitation reaction was performed (Fig 16). C33a cells were untransfected or transfected with an E2 expressing plasmid and lysates were immunoprecipitated with E2-specific antibody (mouse TVG 261). CTCF expression is observed in both untransfected and transfected cell lysates since it is an endogenous protein. E2 expression was confirmed in the transfected cells (fig 16 lane 2) through the presence of a band corresponding its molecular weight. While immunoprecipitation of untransfected lysates with E2 specific antibody did not result in co-precipitation of CTCF (fig 16 lane 3), immunoprecipitation of lysates which contained E2 protein resulted in a clear co-precipitation of CTCF protein, indicating that these proteins exist in a complex (fig 16 lane 4). As an additional control for this experiment, E2-expressing cell lysate was co-immunoprecipitated with non-specific mouse IgG. No visible bands are observed indicating that the co-immunoprecipitation of CTCF with E2 is specific (fig 16 lane 5). E2 transfected lysates was co-immunoprecipitated with monoclonal TVG 261 antibody, a band corresponding the molecular weight of E2 was observed however no visible band were detected for CTCF despite repeated attempts (fig 16 lane 6).
The aim of this experiment was to determine whether E2 and CTCF proteins exist in a complex and determine whether CTCF could indeed recruit E2 to the HPV genome independent of E2 consensus binding sites. From this co-immunoprecipitation experiment it is evident that CTCF protein associates with E2 protein and that this could be the mechanism by which E2 is recruited to CTCF bound to the HPV16 genome.

Figure 16 Co-immunoprecipitation of CTCF with HPV16 E2. 10% input samples are on the left (lane 1 and 2) and the co-immunoprecipitated samples on the right (lane 3-7). The arrows indicate the stated protein. Membrane A was incubated with CTCF antibody, membrane B was incubated with HPV 16 E2 sheep antibody and membrane C was incubated with β-actin as a loading control. In lane 3 UT cell lysate was IP with TVG 261 (mouse) and WB with HPV16 (sheep) E2, lane 4 E2 transfected cell lysate was IP with CTCF (Rabbit) and WB with HPV16 E2, lane 5 E2 transfected cell lysate was IP with non-specific IgG (mouse) and WB with HPV16 E2, lane 6 E2 transfected cell lysate was IP with TVG 261 and WB with HPV16 E2 and lane 7 UT cell lysate was IP with CTCF antibody and WB with HPV16 E2.
C33a cells were either untransfected or transfected with 5µg of FLAG tagged HPV 16 E2 and lysates were immunoprecipitated with FLAG antibody (mouse). Bands corresponding the molecular weight of HPV 16 E2 were observed in the transfected cells (data not shown), however no distinct band for CTCF was observed. Due to time constraints the experiment could not be repeated to improve the quality of immunoprecipitation.

**Expression of CTCF in tissue sections**

So far I have been investigating where in the genome CTCF and HPV 16 E2 are binding using an *in vitro* model system, however expression of CTCF *in vivo* may provide valuable information on disease progression and development. In order to investigate the expression of CTCF in physiologically normal tonsil tissue (n= 10) and in HPV positive (n = 11) and HPV negative (n = 7) tonsil tumour sections were stained with a CTCF-specific antibody by immunohistochemistry. The methodology of the staining technique had to be rigorously optimised prior to analysis of CTCF expression in the various sections that were stained. Primary antibody was serially diluted for 1:200 to 1:3000 in a stepwise manner and two separate conditions were used for antigen retrieval; citric acid was used at pH 6.0 and at pH 9.0. Subsequent analysis of the sections stained with this range of conditions showed that antigen retrieval at pH 6.0 and antibody diluted to 1:2000 was optimal for CTCF staining.

In the normal tonsil section, CTCF expression was detected in majority of the areas. However expression of CTCF in the germinal and follicular area were stronger compared with other areas of the normal tonsil section (Fig 17 A). Positive staining with CTCF was observed for the reticular crypt epithelium as well as the salivary glands. Staining was generally much stronger in the basal cells of the epithelium with loss of expression as the cells became differentiated in the upper layers of the epithelium (Fig 17 B). However, in some areas cells in the basal layer did not stain
for CTCF although cells in the super basal layer stained positive for CTCF, indicating that the loss of expression in the basal layers was due to an artefact of staining caused by incomplete coverage with the antibody. Adipocytes and the sub-mucosal areas of the tonsil were negative for CTCF expression (data not shown).

In the HPV positive tonsil sections, all tumour areas strongly stained positive for CTCF, with strong staining observed in the nuclei (Fig 17 D, E and F). In these sections, weak staining was also observed in the cytoplasm (Fig 17 G). Infiltrating lymphocytes were negative for CTCF expression. Certain areas within the HPV 16 positive tonsil section appeared morphologically normal these areas stained negative for CTCF. In some sections high-grade dysplasia was observed with microinvasion. These areas stained strongly for CTCF (Fig 17 C).

With HPV negative tonsil section, CTCF expression was observed at a lower intensity compared to HPV 16 positive sections. CTCF expression between normal and tumour regions was observed at the same intensity within the HPV negative sections (Fig 17 H). CTCF staining revealed cells forming tumour islands and there were cells that were exhibiting dysplastic characteristics in HPV negative sections.
Figure 17 Representative photomicrographs showing immunohistochemical analysis of CTCF expression. (A) Germinal centre of the normal tonsil exhibits strong CTCF expression in comparison to other areas within the normal tonsil section. (B) Surface epithelium of the normal tonsil showing strong CTCF expression in the undifferentiated basal cells and progressive loss of CTCF expression as the cells become more differentiated. (C) Dysplastic epithelium in HPV negative section showing strong expression of CTCF throughout the epithelium. (D) and (F) HPV 16 positive sections illustrating invasive squamous cell carcinoma of the oropharynx showing strong CTCF expression in the epithelium and loss of differentiation. (E) HPV 16 positive section showing a loss of CTCF expression as the cells migrated from the basal layer up towards the epithelial layer in the pathologically normal regions of the section. (G) HPV positive invasive squamous cell carcinoma of the oropharynx in detail (X 60) showing strongly stained neoplastic cells with pleomorphic nuclei. (H) HPV negative invasive squamous cell carcinoma of the oropharynx with weaker staining of CTCF in comparison to HPV positive equivalents. (I) Negative control (no primary antibody). Magnification A-F, H and I at X20 and G at X60 • symbol represents hyperchromatic nuclei
**DISCUSSION**

HPV 16 E2 maintains the viral life cycle by controlling viral gene replication and transcription\textsuperscript{149,150}. HPV16 E2 is a sequence specific protein, which binds to its consensus sequence and initiates transcription and replication of the viral genome \textsuperscript{1}. In high-risk HPV genomes, there are four highly conserved E2 binding sites located in the viral locus control region\textsuperscript{148}. However there are alternative binding sites, which lie outside the viral LCR and are crucial in the regulation of viral gene expression.

The first part of my project was focused on in the identification of novel HPV 16 E2 binding sites in vivo. Using a ChIP technique, binding of E2 was identified at around 4500 bp and at 6000 bp corresponding to the late gene region of the HPV16 genome and a region where there are no consensus binding sites for E2 (Fig 14). My own bioinformatic analysis has revealed that several non-consensus E2 binding sites are present in this area of the genome, but it is not known whether E2 can bind to these sequences. For example the sequence ACC(N)\textsubscript{7}GT is present at positions (35, 59, 499, 3786, 4591, 5117, 5872, 7019, 7451 and 7858) and AC(N)\textsubscript{7}GGT is present at positions (45, 59, 4418, 5840, 5921, 6243, 6516, 7451 and 7860). It is possible that E2 binds to these non-consensus sites within the late gene region directly, or that E2 is recruited to alternative regions of the genome by other factors that bind to the DNA and recruit E2 via a protein-protein interaction.

Interestingly, only the commercially available HPV 16 E2 antibody was successful in immunoprecipitating E2-bound DNA within the early region of the genome. However, binding in the late regions of the genome was not detected with this antibody. The commercially available HPV 16 E2 antibody is a monoclonal antibody that recognises and binds to a specific epitope in the N-terminal domain of HPV16 E2. In contrast, the polyclonal sheep antibody made in the Parish lab was capable of precipitating E2-bound DNA from the late region, with an apparent lower affinity for
E2 bound to the LCR. It is possible that the epitope for the monoclonal antibody is masked by the binding of an alternative molecule to E2, which inhibits binding of the antibody to its complementary antigen. The polyclonal antibody is presumably not affected by this as numerous epitopes for this antibody are likely to exist. In addition, the E2 binding sequences outside of the LCR may vary leading to a lower binding affinity that is disrupted by binding of the monoclonal antibody.

The ChIP experiments were repeated using chromatin sheared via sonication method rather than enzymatic shearing in order to increase efficiency of the ChIP. In these experiments, the binding of HPV 16 E2 was detected in the LCR and at around 4400, 5600 and 7200 bp with the polyclonal sheep E2 antibody (Fig 15). Previously, binding in the early region was only detected with the commercially available E2 antibody and immunoprecipitation with the HPV 16 E2 antibody designed by the Parish lab was not detected in this region. However with the sonicated chromatin, two E2 binding sites were observed within the LCR, which were not observed in the previous ChIP experiment using enzymatically digested chromatin. Use of the sonication method to shear the chromatin could be contributing towards the detection of additional HPV16 E2 binding sites. With the sonication method, chromatin is sheared more efficiently in comparison to enzymatic shearing, which only digests between nucleosomes. Efficient shearing of the chromatin could enable more sensitive detection of binding above a threshold level.

Put together, the ChIP analyses show that the majority of the E2 binding were observed outside the LCR, this outcome was unexpected since there are four known E2 binding sites within the LCR. It is possible that the E2 binding sites within the LCR are subjected to methylation in the HPV16 genome containing tonsil keratinocytes used for these experiments. The majority of E2 binding sites contain a minimum of one CpG dinucleotide and methylation of the E2 binding site abrogates E2 binding and transcription regulation. It has been reported that
there are five CpG islands within the HPV16 LCR\textsuperscript{153} and that the E2 binding sites are heavily methylated in SCC, with an increase in methylation was detected within E2 BS1 and E2 BS2\textsuperscript{154}. Methylation of the E2 binding sites could correlate to an increased E6 and E7 expression. Methylation in the culture system would give the primary human tonsil keratinocyte cells a proliferative advantage as presumably they would express higher levels of E6 and E7. Since E6 and E7 are oncoproteins encoded from the viral genome\textsuperscript{155}, this would drive the cell towards becoming carcinogenic. Whether methylation of E2 sites in the LCR promotes binding to cryptic sites in the late region is an interesting question that should be addressed.

Recent studies performed by Johansson \textit{et al} revealed inhibition of the early polyadenylation signal (pAE), positioned at nucleotide 4215 by high levels of E2, which leads to transcription of the late viral genes\textsuperscript{156}. pA controls the expression of proteins by adding a poly (A) tail on the mRNA, which acts a signal to RNA polymerase II indicating termination of transcription\textsuperscript{157}. One of the novel E2 binding sites detected in this project was around 4500 bp and therefore located just downstream of the pAE site. E2 was shown to bind to the region of DNA around the pAE between nucleotides 4100 and 4400 and the authors speculated that this could lead to the inactivation of pAE signal thus promoting the expression of late viral genome as shown by\textsuperscript{156}. Further analysis suggested the E2 was recruited by the pAE by the polyadenylation complex, but it is possible that recruitment is via association with CTCF which has also been shown to bind in vitro to this region of DNA, although the CTCF site identified are located slightly downstream of the pAE site (5119, 6127, 6515 and 6860 bp) (parish, unpublished). This could be tested by mutation of the CTCF binding site in this region and analysis of late gene expression in differentiating epithelium.
Interestingly binding of CTCF has been observed in the early region of the genome in cervical cancer and in foreskin keratinocytes. However, with the primary human tonsil keratinocytes CTCF binding was observed in the late region of the genome.

In order to determine where in the host genome CTCF and HPV 16 E2 are binding, ChIP Seq reactions were constructed with the sonicated chromatin. However the final concentration of the ChIP Seq DNA library was insufficient therefore the next generation sequencing reaction was abandoned. Nonetheless this could be further optimised by starting with more concentrated sonicated chromatin, which would provide valuable information on viral genome maintenance mechanisms and pathological status.

From the data analysis it is evident that both HPV 16 E2 and CTCF are binding to the same regions of the HPV16 genome (Fig 14). However it is difficult determine whether these proteins are binding independently to the same region of the HPV16 genomes, or whether they form a complex that co-binds to the DNA. To further establish this interaction, the following co-immunoprecipitation (Co-IP) reaction was constructed. C33a cells were transfected with HPV 16 E2 and lysates were immunoprecipitated with CTCF antibody and co-precipitating E2 was detected. On the other hand immunoprecipitation with E2-specific antibody did not result in Co-IP of CTCF. This could be because the epitope that the E2 antibody recognises is masked when E2 is in complex with CTCF.

From this Co-IP experiment I postulate that CTCF is binding to DNA before associating with E2 protein. This could represent a novel mechanism by which HPV has evolved to recruit E2 to the late region in the absence of E2-specific binding sites. Alternatively CTCF may have a higher affinity for the E2 binding sites, since the binding of E2 varies from sequence to sequence and E2 binding is dependent on flexibility of the target DNA sequence.\textsuperscript{158}
However we have to keep in mind that the interactions I am observing here may not be direct protein-protein interactions. There is a possibility the observed interaction is mediated by a protein complex that recruits CTCF and E2 along with it. The interaction between CTCF and E2 could be mediated by the bridging with a specific DNA sequence. In order to address this question, I could utilise DNase I, which would cleave any DNA that could potentially be bridging a complex between CTCF and E2 or use ethidium bromide, which would intercalate with the DNA and disrupt DNA-protein interactions. Nevertheless, the interaction between CTCF and E2 highlighted in this project is very exciting since it represents a novel mechanism of E2 association with the HPV genome and could further our understanding of the viral life cycle and subsequent implication with disease progression.

CTCF expression \textit{in vivo} may provide valuable information on disease progression and development. In order to investigate the expression of CTCF in physiologically normal tonsil tissue (n= 10) and in HPV positive (n = 11) and HPV negative (n = 7) tonsil tumours, sections were stained with a CTCF-specific antibody by immunohistochemistry. Staining conditions were optimised to best illustrate the expression of CTCF in normal tonsil tissue and in HPV positive and HPV negative tonsil tissue. Antibody was serially diluted until background staining was absent, a dilution of 1:2000 provided good specificity and sensitivity. There are two commercially available antigen retrieval solutions one of which is at pH 6 while the other is at pH 9. Both pH 6 and 9 were used during the antigen retrieval process however antigen retrieval at pH 6 provided better a staining with CTCF.

Immunohistochemistry analysis revealed CTCF expression in majority of the areas of the normal tissue section. Nonetheless an increased expression of CTCF was observed in the germinal centre and follicular areas. In the HPV positive sections, all tumour areas stained strongly for CTCF, with strong staining observed in the nuclei (Fig 17 D and F) and weak staining was observed in the cytoplasm (Fig 17 G).
Regions of high-grade dysplasia as well as micro invasion also stained strongly for CTCF. Staining of cells in the basal layer was strong, indicating a high level of CTCF expression. A gradual decrease in CTCF staining was observed as the cells migrated from the basal layer to the super basal layer and up towards epithelial layer (Fig 17 E). One reason behind this could be the expression of late genes in the differentiated epithelium where CTCF expression is normally reduced. Therefore I postulate that a reduction in CTCF recruitment to the late region of the HPV genome is essential for the induction of late gene expression and viral infection persistence. Recent studies performed by Garrido et al also showed a decrease in CTCF expression in the epithelial layer. However, their results are not entirely convincing and the images presented in this project provide clear evidence that CTCF expression is switched off with epithelial cell differentiation. In addition intense CTCF staining was observed with the tumour area (Fig 17 E).

In the HPV negative sections, expression of CTCF was generally lower in comparison to expression in the HPV positive sections. Interestingly, CTCF expression between normal and tumour regions was observed at the same intensity within the HPV negative sections. CTCF staining revealed cells forming tumour islands and there were cells that were exhibiting dysplastic characteristics in HPV negative sections.

It is interesting to see a difference in CTCF expression between HPV related cancer and non-HPV related cancer. This could be an indication that CTCF plays a major role in HPV-related cancer by facilitating early gene expression and maintenance of the genome and that HPV infection results in a specific up-regulation of CTCF to support the virus life cycle. However more analysis is needed to further establish this effect of CTCF staining.
CONCLUSIONS

The main goal of my investigation was to identify novel E2 binding sites outside the LCR and determine their role in the regulation of viral gene expression. Bioinformatic analysis revealed several non-consensus E2 binding sites downstream to the LCR, nevertheless binding of E2 to these sites are not well characterised. In order to identify novel E2 binding sites outside the LCR primary human tonsil keratinocytes containing HPV16 genome were used to perform ChIP. Which revealed E2 binding sites at around 4400, 4500, 5600 and at 6000 bp this corresponds to the late gene region of the HPV16 genome and in this region there are no consensus E2 binding sites. In my opinion this could be due the methylation of E2 binding site within the LCR and methylation of these sites prevents binding of E2 and subsequently abolishes its functions. Methylation could promote the expression of E6 and E7 viral oncoproteins and these proteins would drive the cell towards tumorigenesis. However the presence of methylation at these sites needs to be investigated, with the aid of pyrosequencing.

One of the E2 binding sites identified was located upstream to the pAE site and it is shown that binding of E2 to this site leads to the expression of late viral genome. It is possible that E2 is recruited to the pAE site with the help of CTCF, which was observed to bind at this site. To further establish this CTCF binding site at the region could be mutated and the late genome expression could be analysed.

From the ChIP analysis it was evident that E2 and CTCF are binding in the same region of the HPV16 genome. To further establish this, a Co-IP reaction was constructed and after analysing the data I believe that CTCF is binding to DNA before associating with E2 protein. Alternatively CTCF may have a higher affinity to the E2 binding site. To further evaluate this DNase I could be utilised to eliminate
any potential DNA bridging complex or use ethidium bromide to intercalate DNA and disrupt DNA-protein interactions.

CTCF expression was investigated on normal, HPV positive and HPV negative tonsil section. In the normal section increased expression of CTCF was observed in the germinal centre and follicular area. In the HPV positive section strong nuclei staining and weak cytoplasmic staining was observed. As well as high grade dysplasia and microinvasion. CTCF expression was higher in the basal layer however expression of CTCF was switched off in the differentiated epithelial layer. HPV negative section stained at a lower intensity for CTCF in comparison to HPV positive section. Nonetheless more tissue section needs to be stained with an alternative antigen retrieval method such as PIER method. The PIER method uses enzymes such as Proteinase K, Trypsin, and Pepsin, these enzymes are able to cleave peptides which are masking the epitope.
<table>
<thead>
<tr>
<th>Primer Pair</th>
<th>Sequence</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2 Fw</td>
<td>CAGGGTGCGGTACAGGCGGA</td>
<td>1.9347</td>
</tr>
<tr>
<td>L2 Rv</td>
<td>GGATCGGAAGGGCCCCACAGGA</td>
<td></td>
</tr>
<tr>
<td>L2 CTCTbs Fw</td>
<td>AGGCCTACTGGCATAGGTACAGT</td>
<td>1.7444</td>
</tr>
<tr>
<td>L2 CTCTbs Rv</td>
<td>AGGTAAAGGCTGCTGTAAGGTGCTGT</td>
<td></td>
</tr>
<tr>
<td>L2/L1 Fw</td>
<td>TGGCTGCGCTAGAGGCACCCTGT</td>
<td>1.9</td>
</tr>
<tr>
<td>L2/L1 Rv</td>
<td>TGGCTGCAACATATTCATCCCGTGC</td>
<td></td>
</tr>
<tr>
<td>L1 CTCFbs Fw</td>
<td>TGCAGCAAATGCAAGTGGTGAGT</td>
<td>1.8647</td>
</tr>
<tr>
<td>L1 CTCFbs Rv</td>
<td>TGGGGATCTCCTTGCCCCAGTGT</td>
<td></td>
</tr>
<tr>
<td>L1 Fw</td>
<td>ACAAGCAGATGGAAGCGCAAACCA</td>
<td>1.9379</td>
</tr>
<tr>
<td>L1 Rv</td>
<td>AGAGTGATAGTGAGTGGGGTGGT</td>
<td></td>
</tr>
<tr>
<td>5’URR-Enh Fw</td>
<td>TTTTGATAGCCACCGCGCCATT</td>
<td>1.8501</td>
</tr>
<tr>
<td>5’URR-Enh Rv</td>
<td>GCATGGCCAGCAACGCCAGTACA</td>
<td></td>
</tr>
<tr>
<td>Enh Fw</td>
<td>CCAAATCCCTGTCTTTCTCTGA</td>
<td>1.6585</td>
</tr>
<tr>
<td>Enh Rv</td>
<td>GTTGGCCGCATAGTATTTA</td>
<td>1.5128</td>
</tr>
<tr>
<td>Earlyprom Fw</td>
<td>GCAAACCGTTTGGGTGTACA</td>
<td></td>
</tr>
<tr>
<td>Earlyprom Rv</td>
<td>ACTAACCCTTCCCCTGCCTCA</td>
<td></td>
</tr>
<tr>
<td>E6 111-223 Fw</td>
<td>AGGACCCACAGGACCGACCC</td>
<td>1.8695</td>
</tr>
<tr>
<td>E6 111-223 Rv</td>
<td>AGTCGCGATACCTGGCTGTGCCA</td>
<td></td>
</tr>
<tr>
<td>E6 427-506 Fw</td>
<td>GCCACTGTGTCTGTAGAAGAAAAGCA</td>
<td>1.9332</td>
</tr>
<tr>
<td>E6 427-506 Rv</td>
<td>GCCGGTCCACCGACCCCTTT</td>
<td></td>
</tr>
<tr>
<td>Lateprom Fw</td>
<td>GACAGCTCAGAGGAGAGGAGGA</td>
<td>1.9004</td>
</tr>
<tr>
<td>Lateprom Rv</td>
<td>GCACAACCACCAAGCCTAGAGT</td>
<td></td>
</tr>
<tr>
<td>E1 1250-1368 Fw</td>
<td>CGCAGACAGCGGGTAGTGGCA</td>
<td>1.9154</td>
</tr>
<tr>
<td>E1 1250-1368 Rv</td>
<td>GCAAACCCCCACCTCCACCC</td>
<td></td>
</tr>
<tr>
<td>E1 2158-2316 Fw</td>
<td>AGGGTAGATGATGGAGGTGTGG</td>
<td>1.8461</td>
</tr>
<tr>
<td>E1 2158-2316 Rv</td>
<td>GATTACCTCTGTGTAGTGCACCA</td>
<td></td>
</tr>
<tr>
<td>E2 (CTCF) Fw</td>
<td>GGAACACATAGGGCCTAGAATGTGC</td>
<td>1.8923</td>
</tr>
<tr>
<td>E2 (CTCF) Rv</td>
<td>TGATACAGCCAGTGTTGGCAACC</td>
<td></td>
</tr>
<tr>
<td>E4 Fw</td>
<td>CACTCCGCGCGCGACCATAC</td>
<td>1.9326</td>
</tr>
<tr>
<td>E4 Rv</td>
<td>GGGTGGCGAGGGGTTTCCCGG</td>
<td></td>
</tr>
<tr>
<td>E5 Fw</td>
<td>AGTCCGCTCGCTTTGTGTGGT</td>
<td>1.9129</td>
</tr>
</tbody>
</table>
Table 4 represents the primers used during qPCR, which cover the entire HPV16 genome. The primers were diluted to a final concentration of 10 pmol ml⁻¹.

<table>
<thead>
<tr>
<th>NTB R</th>
<th>Primer Product</th>
<th>NEBNext Index Primer Sequence</th>
<th>Expected Index Primer Sequence Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE7330A: 0.010 ml</td>
<td>E7330A: 0.020 ml</td>
<td>NEBNext Index 1 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
<tr>
<td>HE7410A: 0.010 ml</td>
<td>E7410A: 0.020 ml</td>
<td>NEBNext Index 2 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
<tr>
<td>HE7411A: 0.010 ml</td>
<td>E7411A: 0.020 ml</td>
<td>NEBNext Index 3 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
<tr>
<td>HE7442A: 0.010 ml</td>
<td>E7442A: 0.020 ml</td>
<td>NEBNext Index 4 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
<tr>
<td>HE7444A: 0.010 ml</td>
<td>E7444A: 0.020 ml</td>
<td>NEBNext Index 5 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
<tr>
<td>HE7445A: 0.010 ml</td>
<td>E7445A: 0.020 ml</td>
<td>NEBNext Index 6 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
<tr>
<td>HE7446A: 0.010 ml</td>
<td>E7446A: 0.020 ml</td>
<td>NEBNext Index 7 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
<tr>
<td>HE7448A: 0.010 ml</td>
<td>E7448A: 0.020 ml</td>
<td>NEBNext Index 8 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
<tr>
<td>HE7449A: 0.010 ml</td>
<td>E7449A: 0.020 ml</td>
<td>NEBNext Index 9 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
<tr>
<td>HE7450A: 0.010 ml</td>
<td>E7450A: 0.020 ml</td>
<td>NEBNext Index 10 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
<tr>
<td>HE7451A: 0.010 ml</td>
<td>E7451A: 0.020 ml</td>
<td>NEBNext Index 11 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
<tr>
<td>HE7452A: 0.010 ml</td>
<td>E7452A: 0.020 ml</td>
<td>NEBNext Index 12 Primer for Illumina</td>
<td>5’-CAAGCAGAGAGCCCGCCGAGATTCGCTCTCCGAGATG-3’</td>
</tr>
</tbody>
</table>
References


61. Jones, D. L., Thompson, D. A. & Munger, K. Destabilization of the RB tumor suppressor protein and stabilization of p53 contribute to HPV type 16 E7-


106. Holmgren, S. C., Patterson, N. A., Ozbun, M. A. & Lambert, P. F. The minor capsid protein L2 contributes to two steps in the human papillomavirus type 31


137. El-Kady, A. & Klenova, E. Regulation of the transcription factor, CTCF, by


Binding Site 1 (E2BS1), E2BS2, and the Sp1-Binding Site in Cervical Cancer Samples as Determined by High-Resolution Melting Analysis-PCR. *Journal of Clinical Microbiology* (2013) 51, 3207–3215.


