

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

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

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 Siddiqi 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 Zeidler 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

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

MATERIALS AND METHODS	28
Chromatin Immunoprecipitation (ChIP)	28
Chromatin preparation	28
Enzymatic Shearing of Chromatin	28
Chromatin shearing by sonication	29
Immunoprecipitation of Chromatin	29
Chromatin elution	30
Real-time PCR	31
Library preparation for ChIP-SEQ	32
Cell culture	32
Cell lysis	33
Co-Immunoprecipitation	33
Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting	34
Detection of the membrane	35
Immunohistochemistry	35
RESULTS	38
Assessment of E2 association with the HPV 16 genome by ChIP	38
End repaired ChIP DNA	44
Co-Immunoprecipitation of C33a cells transfected with HPV 16 E2	45
Expression of CTCF in tissue sections	47
DISCUSSION	52
CONCLUSIONS	58
APPENDIX	60
REFERENCES	62

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)

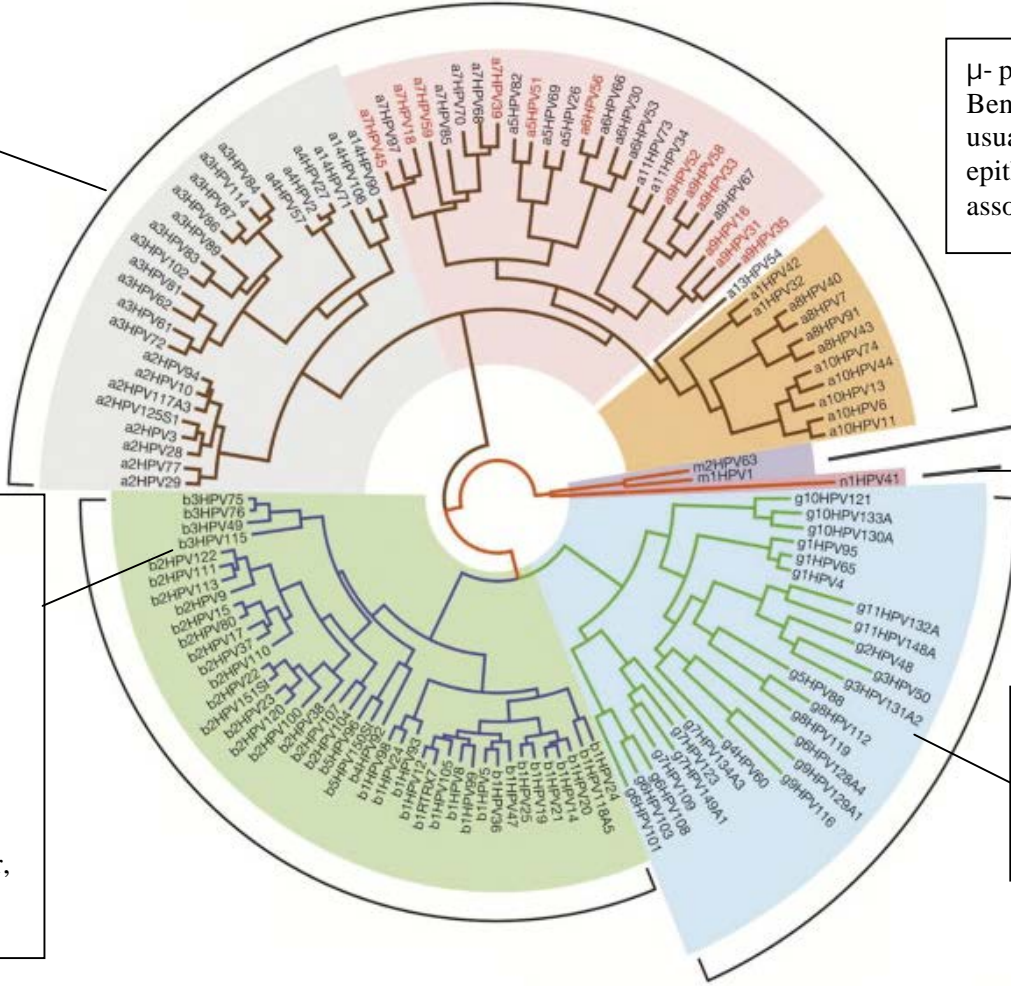
INTRODUCTION

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 α , β , γ , μ and ν 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. β and γ PVs genera only seems to cause asymptomatic infections within immunocompetent individuals ⁹. All mucosal HPV types belong to the α 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, α 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}.

α -papillomavirus mucosal & cutaneous: Group includes high-risk mucosal types that are associated with cervical cancer, low-risk mucosal types that are associated with benign lesions, and low-risk cutaneous types that typically cause skin warts some types can be found in both mucosal and cutaneous lesions, although a preference is usually apparent. Low-risk HPV DNA is only rarely found in cancers. High-risk HPV DNA is found in almost all cases of cervical cancer.



μ -papillomavirus cutaneous: Benign cutaneous lesions usually at palmer and planter epithelial sites. Not associated with cancer

ν -papillomavirus cutaneous: Benign cutaneous lesions. DNA occasionally detected in skin cancer

β -papillomavirus cutaneous lesions: group includes high-risk and low-risk cutaneous types. Typically associated with unapparent infections in the immunocompetent host, but can proliferate in immunosuppressed hosts and in epidermodysplasia verruciformis patients. Persistent infection is thought to predispose to the development of skin cancer, especially in immunosuppressed individuals

γ -papillomavirus: benign cutaneous lesion. Some types detected at oral sites. DNA only very rarely found in skin cancer

Figure 1 illustrates the phylogenetic tree of HPV. HPVs can be divided into five evolutionary groups with various epithelial tropisms and disease association. α PVs are comprised of low-risk mucosal types (orange shaded segment) known to cause benign warts and high-risk mucosal types associated 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 ¹⁴. High-risk HPV types 16 and 18 have been identified as the predisposing factors in 70% of the cases of cervical cancer ¹⁵. 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 ¹⁵. 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).

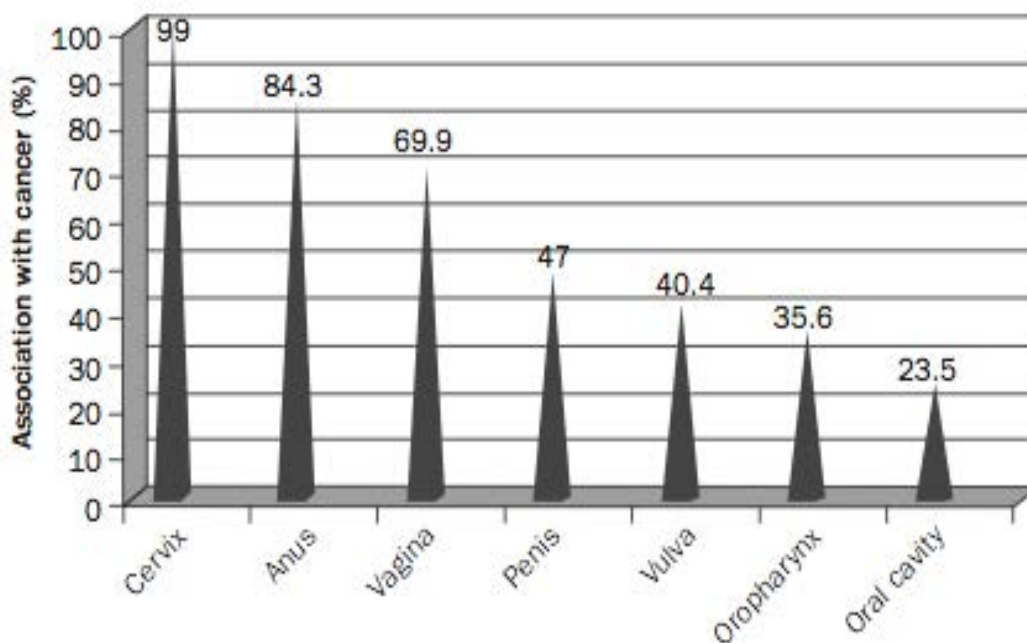


Figure 2 represents the range of cancers associated with HPV. Image taken from ⁷

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 ¹⁶. The prevalence of cervical cancer is higher in developing countries accounting for 15% out of all types of cancers ¹⁷. 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 ¹⁸. 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 ¹⁷ for example in Kenya, Zimbabwe and Mozambique at 33.6% ¹⁹. In the Caribbean regions such as Trinidad and Tobago 35.4% and in Latin American regions 16.1% prevalence are observed ¹⁹. (Fig 3) ⁷ 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 ¹⁹.

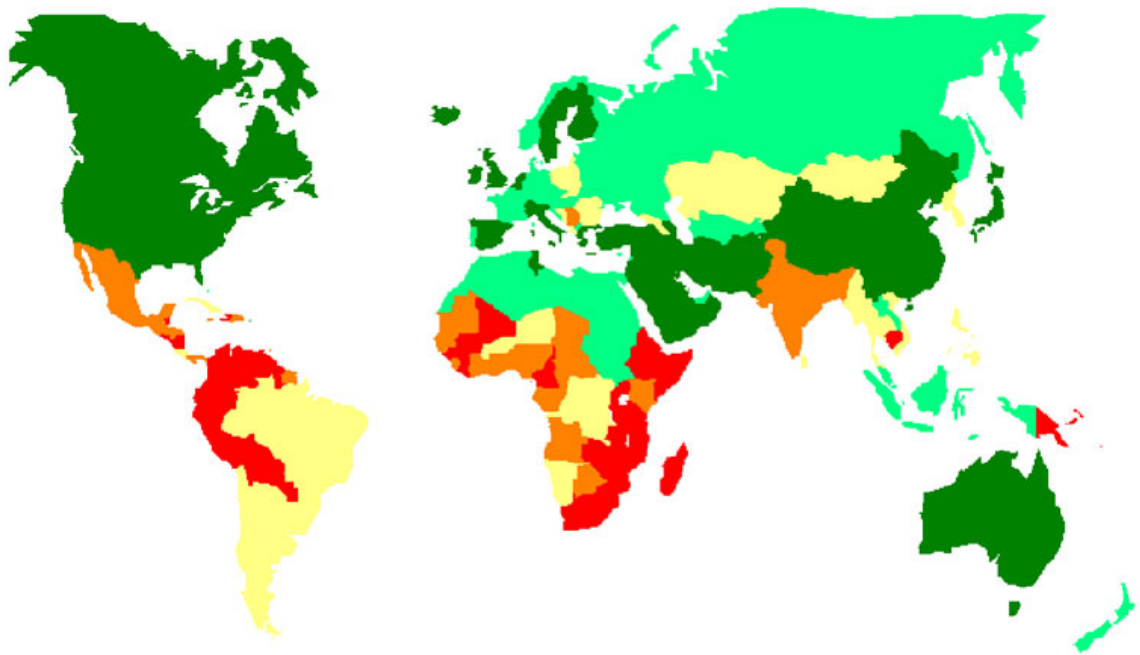


Figure 3 is a representation of HPV prevalence worldwide. Red coloured regions correspond to high HPV prevalence, yellow coloured region correspond to intermediate HPV prevalence and green coloured region correspond to low HPV prevalence ⁷

Prophylactic action against papillomavirus

Cervical screening programmes have significantly reduced the incidence of invasive cervical cancer and the associated mortality rate ²⁰. 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 ¹⁸. 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 ²¹. 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 ²¹.

Vaccination against HPV

Vaccines against HPV were the first generated anti-cancer vaccine ²². 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 ^{23,24}. These vaccines initiate the host immune response mediated via antibodies against the viral protein L1 ²⁵. Administration of Cervarix has shown to provide antibody responses up to 6.4 years ^{26,27} on the other hand the quadrivalent vaccine has been shown to provide effective antibody responses for 5 years ²⁸. However, additional preventive measures such as a screening program, life style changes and physical examination should not be abandoned in vaccinated populations ²² as results from a phase 3 trial reveal the inadequacy of the vaccines

towards existing HPV infection and its progression towards pathological conditions²⁹.

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 region³⁰. 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 DNA²². 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 particles³¹.

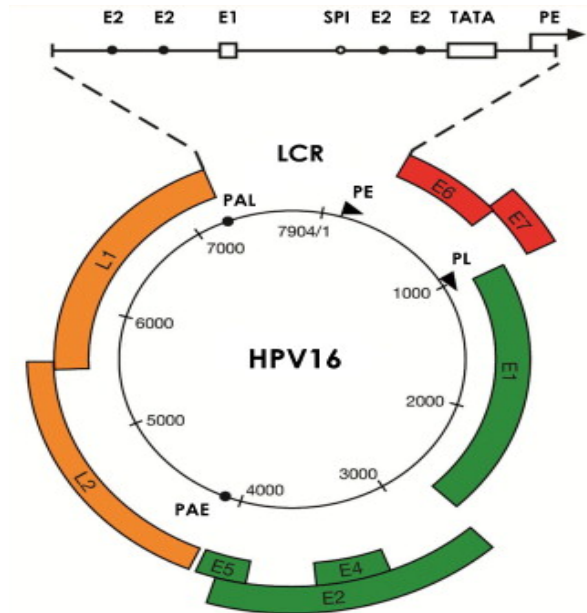


Figure 4 is an illustration of the HPV16 genome. The papillomavirus genome is a circular double-stranded molecule of 8,000 base pairs. There are three distinctive regions within the genome they are the LCR, the early (E) gene region, and the late (L) gene region. The genome encodes essential proteins 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^{34,35} 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^{35,36}. Expression of E6 and E7 are down regulated by viral transcription factor E2^{35,37}. Basal differentiation causes the activation of late viral functions, which are mediated by E4, E5 and E1^{38,39}. Viral DNA replication and protein expression are amplified upon late promoter activation, this leads to an increase in viral copy number⁴⁰.

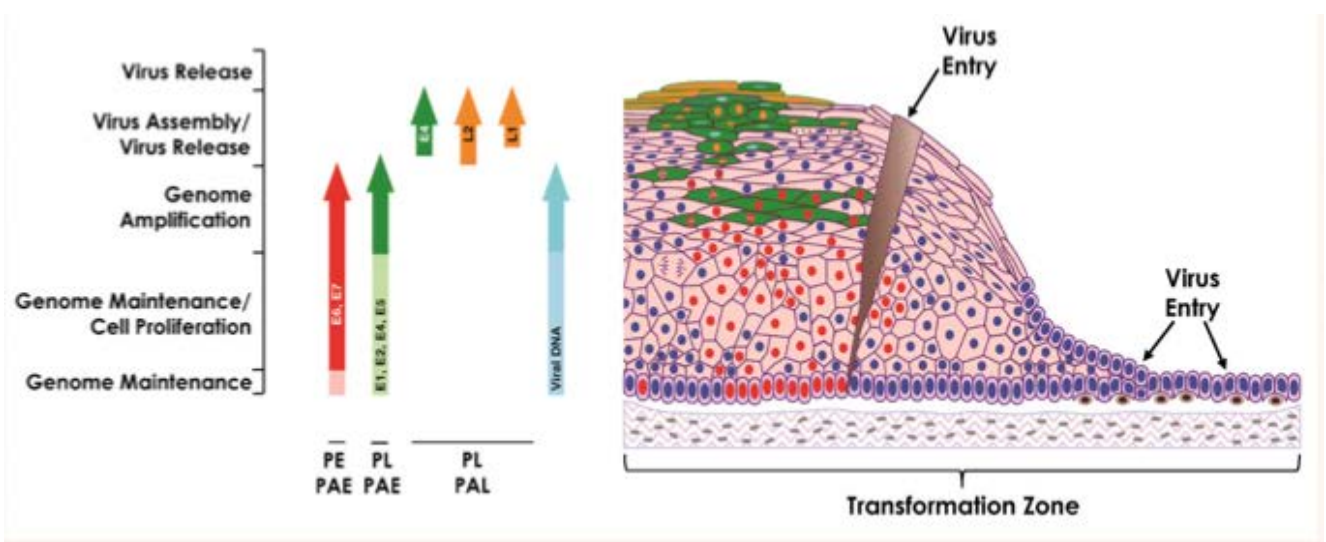


Figure 5 illustrates the viral life cycle in cervical epithelium. The HPV life cycle initiates following infection of the epithelial cells in the basal layer. Cells with red nuclei represent virally infected basal epithelial cells that are dividing. Viral protein E4 expression is up regulated in cells 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 epithelium⁴

Persistent infection drives cell proliferation in the basal and the parabasal cell layers^{41,42}. Infected cells within the basal cell layer undergo mitosis where daughter cells are produced, then migrate towards the epithelial surface and begin to differentiate⁴. 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

genome⁴. 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^{4,22}.

Viral oncoproteins E6 and E7

The oncoproteins E6 and E7 are the main causative agents in the development of HPV induced cervical cancer^{35,37}. E6 consists of two zinc finger motifs and remains localised in the cytoplasm as well as the nucleus of the infected keratinocytes^{43,44}. 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 tumour suppressor protein becomes activated in response to cellular stress and DNA damage response^{45,46}. p53 degradation is initiated by E6 and the E6-associated protein (E6-AP) which target p53 for proteasome-dependent degradation by conjugation of ubiquitin⁴⁷⁻⁴⁹. Degradation of p53 eliminates restraints on DNA synthesis enabling replication of viral DNA^{49,50}. In addition, E6 has the capacity of upregulate and sustain telomerase activity over numerous cell divisions^{51,52} and it has recently been established that E6 degrades PDZ substrates and contributes to tight junction ablation via zona occludens 1 (ZO-1) relocalisation⁵³. PDZ motifs are associated with cell signalling, polarity and proliferation⁵⁴.

E7 consists of three conserved regions and the conserved region 2 contains a LXCXE motif^{55,56}. This motif mediates the binding of E7 to the tumour suppressor protein pRb⁵⁷. pRb is hypophosphorylated thus enabling the inhibition of the transcription factor E2F during the G₁ phase of the cell cycle^{58,59}. 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

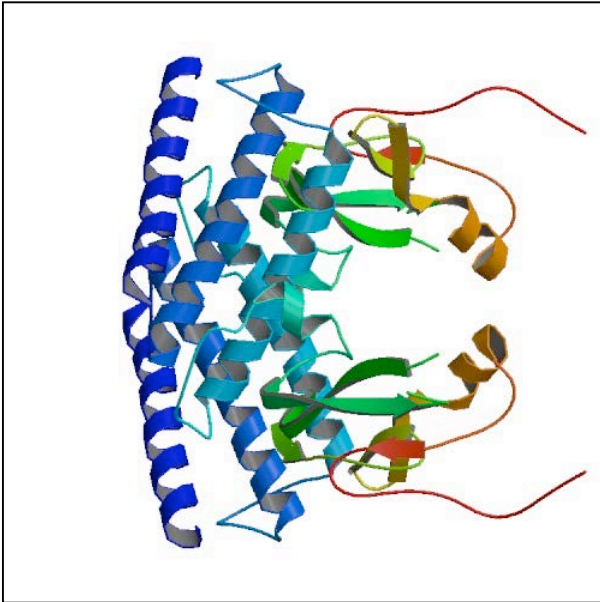
E2 is DNA binding protein encoded by all PVs in the early and intermediate stages of the viral life cycle. E2 regulates HPV genome transcription and viral DNA replication ⁶⁹. DNA binding protein E2 is sequence specific and capable of binding to 12 base pair motifs that are located in the LCR ⁷⁰. The 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 α -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 β -sheets that form an unusual β -barrel structure

surrounded by four α -helices make up the C-terminal DNA binding and dimerization domain ⁷⁵. 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) ⁷⁶.

Transactivation domain



DNA binding domain

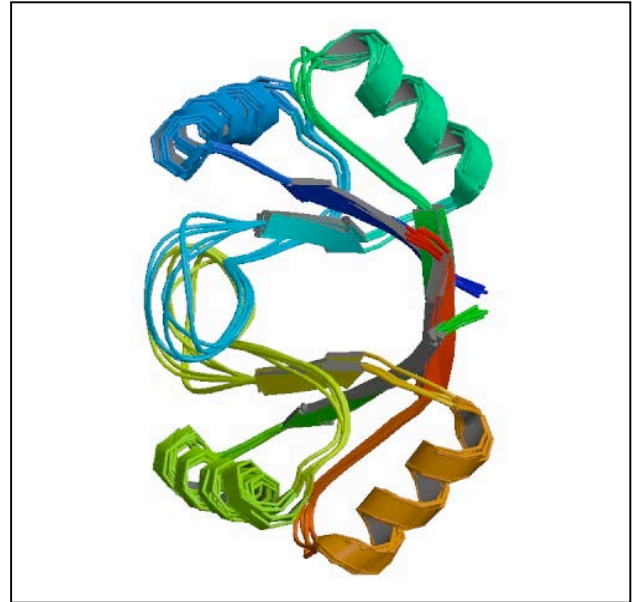


Figure 6 illustrates HPV 16 E2 with transactivation domain with α -helices arranged anti-parallel with each other ⁷⁶ and DNA binding domain with anti-parallel β -sheets are observed ⁷⁵. The two domains are connected via hinge structure

The transactivation domain contains essential residues that orchestrate transcriptional regulation and these residues are located on the outer surface of the domain ⁷³. In contrast, residues that are associated with replication are located on the inner surface of the domain, because they disrupt binding to E1 ⁷². Unlike the transactivation domain sequence specificity does not perturb the DNA binding domain nevertheless it is known to stabilise interactions between DNA and protein ⁷⁷.

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 processes⁷⁸. E2 binds to DNA in a sequence specific manner. There are four well characterised E2 binding sites within the LCR of the α PV genome¹⁰. The binding of E2 to DNA is dependent on the consensus binding sequences they are ACCG(N)₄CGGT or ACC(N)₆GGT 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 repression⁸². 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 α PV species exhibits a weak binding to host chromosomes⁹⁰.

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 helicase⁹¹. Conservation of the E1 protein is indicative of its role as a replicative helicase, which delineates conscientious replication of the viral episome⁹². E1 activity is essential for viral

genome replication, copy number increase and regulation of episomal levels ⁹³. At the viral ori, E1 constructs into a double-hexamer that melts DNA at the origin prior to establishment of the replication fork ^{94,95}.

Viral protein E4

E4 ORF encodes a protein that is variable in size between PV types ⁹⁶. Expression of E4 protein arises from the spliced mRNA product E1[^]E4 that consist of the E1 initiation codon and few sequences from the E1 ORF ^{97,100}. 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 ⁹⁹⁻¹⁰¹, and expression is only detected in differentiated epithelium. E4 has been shown to contribute to keratin network disruption ¹⁰² and which potentially aids towards viral transmission ¹⁰³.

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 ^{4,104}. L2 is the minor capsid protein ¹⁰⁵ and recent studies have established its role in HPV genome encapsidation ¹⁰⁶. 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 ¹⁰⁷⁻¹⁰⁹. L1 is the major viral capsid protein ⁴ 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 cytoplasm¹¹⁰.

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 ¹¹¹. 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 cell¹¹².

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 exist⁷². 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^{3,113}. CCCTC-binding factor (CTCF) is an 11 zinc-finger protein that which 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 nucleosomes¹¹⁴ and it was initially identified as transcriptional repressor for chicken c-myc and lysozyme genes^{115,116}. CTCF has also been shown to function as a transcriptional activator¹¹⁷. 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 activation¹¹⁸⁻¹²⁰. CTCF is associated with a variety of molecular functions depending on the genetic locus involved, such as transcriptional activation^{121,122}, transcriptional repression^{123,124} and enhancer blocking activity^{125,126}. CTCF is involved in a variety of long range processes such as chromatin looping¹²⁷, chromatin insulation¹¹⁴, chromosome segregation¹²⁸, and nuclear organisation^{129,130}. CTCF binding sites are highly conserved across different cell types¹³¹. Interspecies CTCF binding profile within the liver was investigated, which unveiled 5000 highly conserved sites between species and tissues^{132,133}.

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 ^{134,135}. 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 ¹³⁶ (Fig 7). Phosphorylation at the N-terminal region switches CTCF function from transcriptional repressor to transcriptional activator ^{137,138}. PARylation is crucial for CTCF to function as an insulator and the lack of this modification abolishes barrier activity ¹³⁹. SUMO Small ubiquitin like protein is capable of modifying CTCF which enhances the CTCF repressor function at the c-myc P2 promoter ¹³⁴.

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 ¹³². Within certain locations the primary binding motif is capable of defining a CTCF binding site whereas, the secondary binding motif enhances CTCF binding ¹³². Bases that are adjacent to primary and secondary motifs are also thought to be important for CTCF binding ¹⁴⁰.

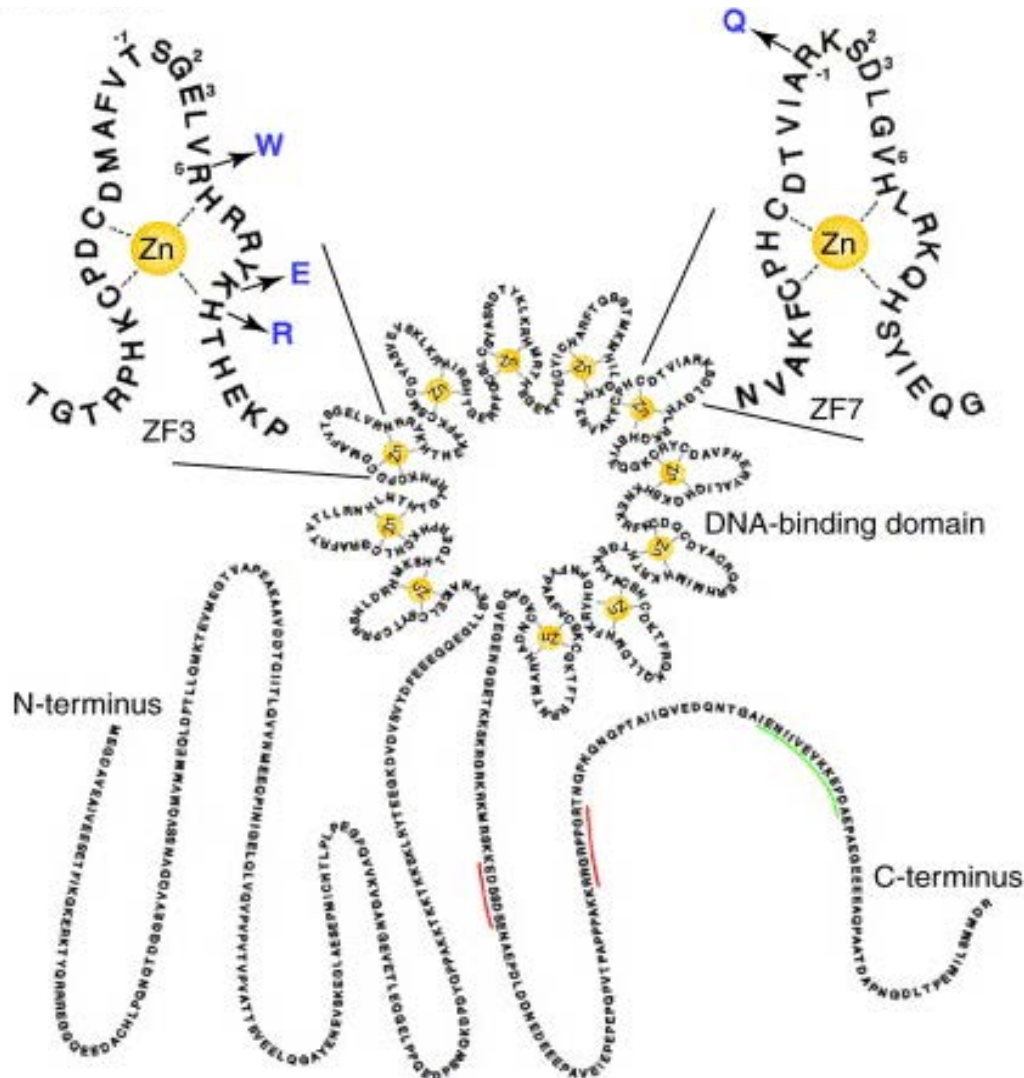


Figure 7 represents CTCF structural features as well as substitution of amino-acid specific to tumours within zinc fingers. CTCF protein present in humans has a DNA-binding domain, which comprises of ten C_2H_2 -class ZFs (ZFs 1–10) and one C_2HC -class ZF (C-terminal ZF11)¹⁶⁰

The 11 zinc fingers of the CTCF DNA binding region do not all commit equally to the binding of DNA¹⁴⁰. Zinc fingers 4-7 target the core DNA binding motif and bind to 80% of the CTCF binding sites, where as zinc fingers 8-11 and 1-2 stabilise CTCF broadly¹²³. 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-ribosylation) 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).

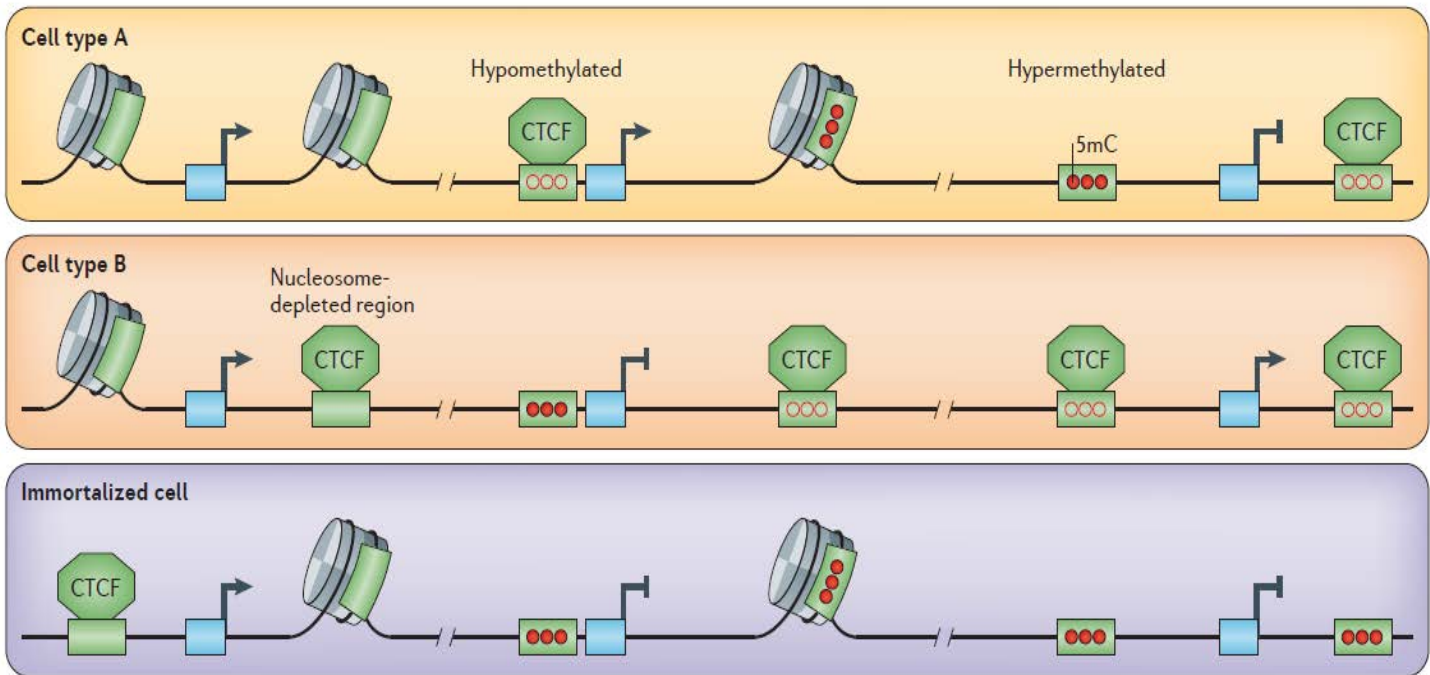


Figure 8 illustrates maintenance of CTCF and DNA binding with each other. Binding of CTCF binding is partially maintained via occupancy of the nucleosomes along with differential DNA methylation at specific CTCF binding sites. This suggests that the cells are able to remodel chromatin complexes by using 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 ¹⁴³

CTCF has a major impact on the three dimensional structure of DNA which in turn has a major impact on gene regulation ¹³⁶. Interaction of CTCF with nucleophosmin/B23 as well as the nuclear matrix, could be the factors contributing towards three dimensional DNA conformation ^{144,145}. 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).

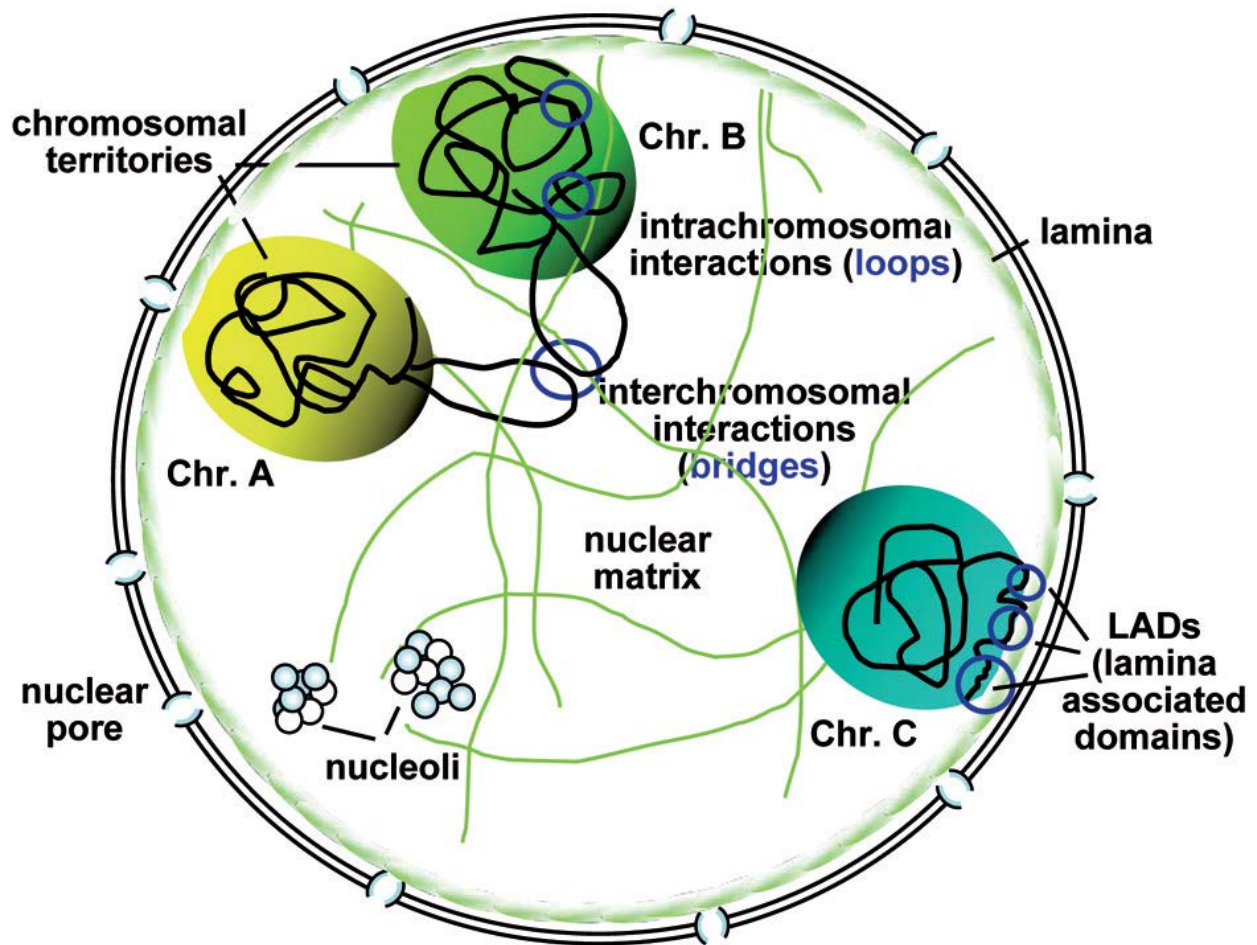


Figure 9 illustrates the positioning and partitioning of DNA mediated by CTCF. Individual interphase chromosomes are formed from Double stranded DNA that occupies a certain area of the nuclear volume, this leads towards the formation of chromosome territories. Within the transcriptionally active centre 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 binding 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 or 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.

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

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.

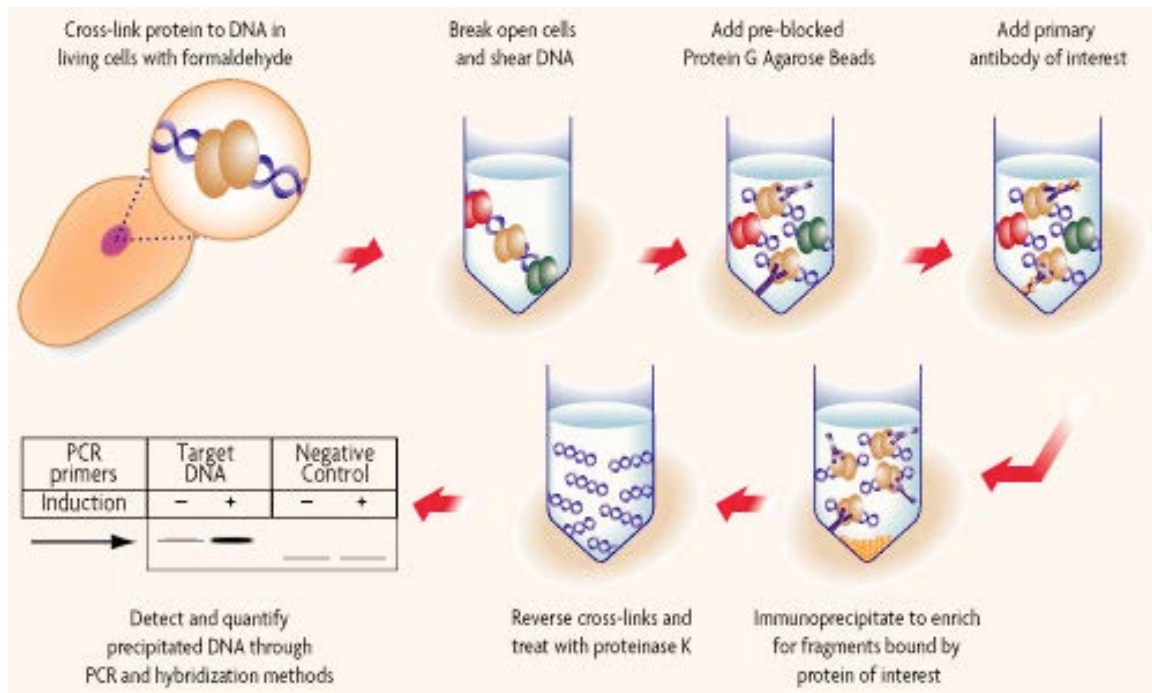


Figure 10 illustrates the workflow of CHIP technique (www.activemotif.com)

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 $10 \text{ pmol} \cdot \mu\text{l}^{-1}$ of which 0.125 ml of forward and 0.125 ml reverse primer were added to each reaction along with 8.75 ml of H_2O and 10 ml of Senimix SYBR No-ROX 2X (Bioline). The final concentration of the primers was $0.42 \text{ pmol} \cdot \mu\text{l}^{-1}$. 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 2×10^6 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.

Antibody Name	Antibody type	Isotype	Manufacturer	Catalogue number	Volume μ l	Species
HPV16 E2	Polyclonal		Parish lab	-	2	Sheep
TVG 261	Monoclonal	IgG1	ABCAM	17185	2	Mouse
IgG	-	IgG	SANTA CRUZ	sc-2025	2	Mouse
CTCF	Polyclonal	IgG	Active Motif	61311/61312	2	Rabbit

Table 2 outlines the antibodies used to perform co-immunoprecipitation

The samples were briefly centrifuged at 6000 x g and the supernatant discarded. The protein G beads were washed three times in 500 μ l 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

5% milk buffer for one hour at room temperature by gently rocking. Table 4 lists the primary antibodies used.

Antibody Name	Antibody type	Isotype	Manufacturer	Catalogue number	Dilution	Species
HPV16 E2	Polyclonal		Parish lab	-	1:500	Sheep
TVG 261	Monoclonal	IgG1	ABCAM	17185	1:1000	Mouse
β -Actin	Monoclonal	IgG1	Sigma-Aldrich	A5441	1:5000	Mouse
CTCF	Polyclonal	IgG	Active Motif	61311/61312	1:1000	Rabbit

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 reprobng 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 region¹⁵¹. 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).

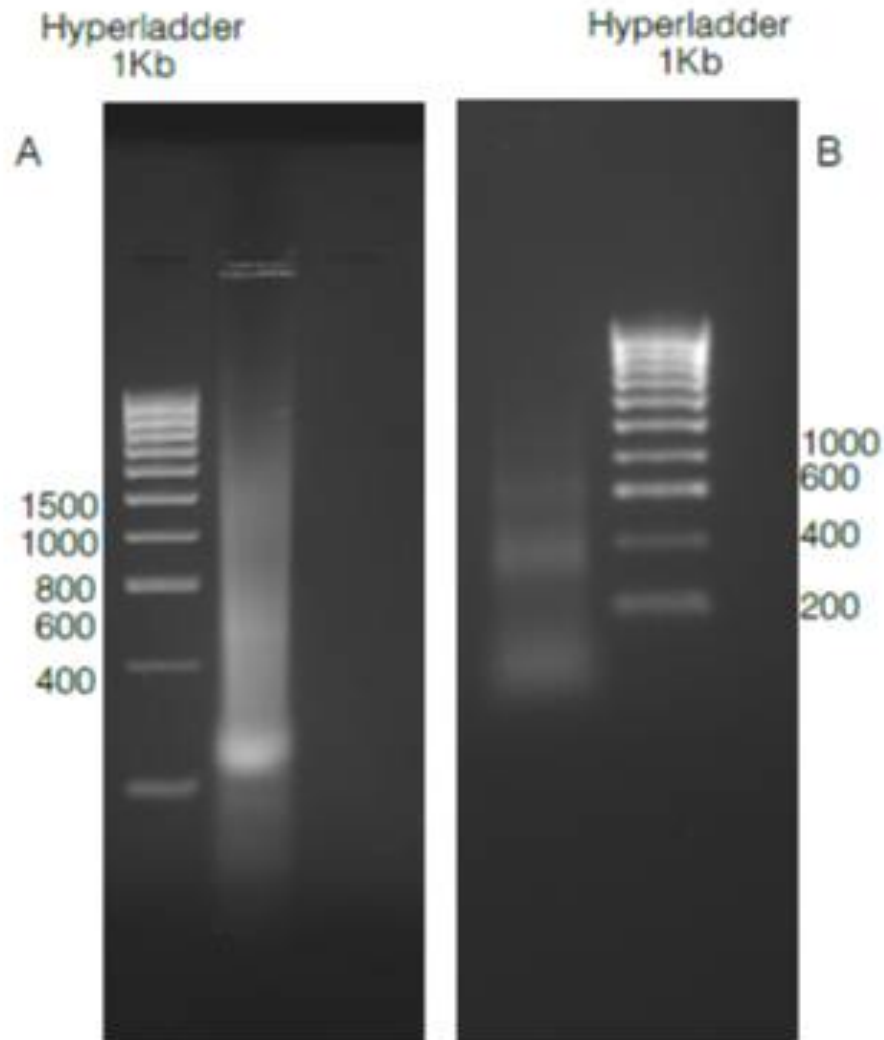


Figure 11. Shearing of the chromatin either via sonication (A) or by enzymatic digestion (B). Sheared chromatin was immunoprecipitated with FLAG antibody, sheep HPV 16 E2 antibody, mouse HPV16 E2 antibody TVG 261 or rabbit CTCF antibody and co-precipitation of the HPV16 genomes analysed using real-time 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.

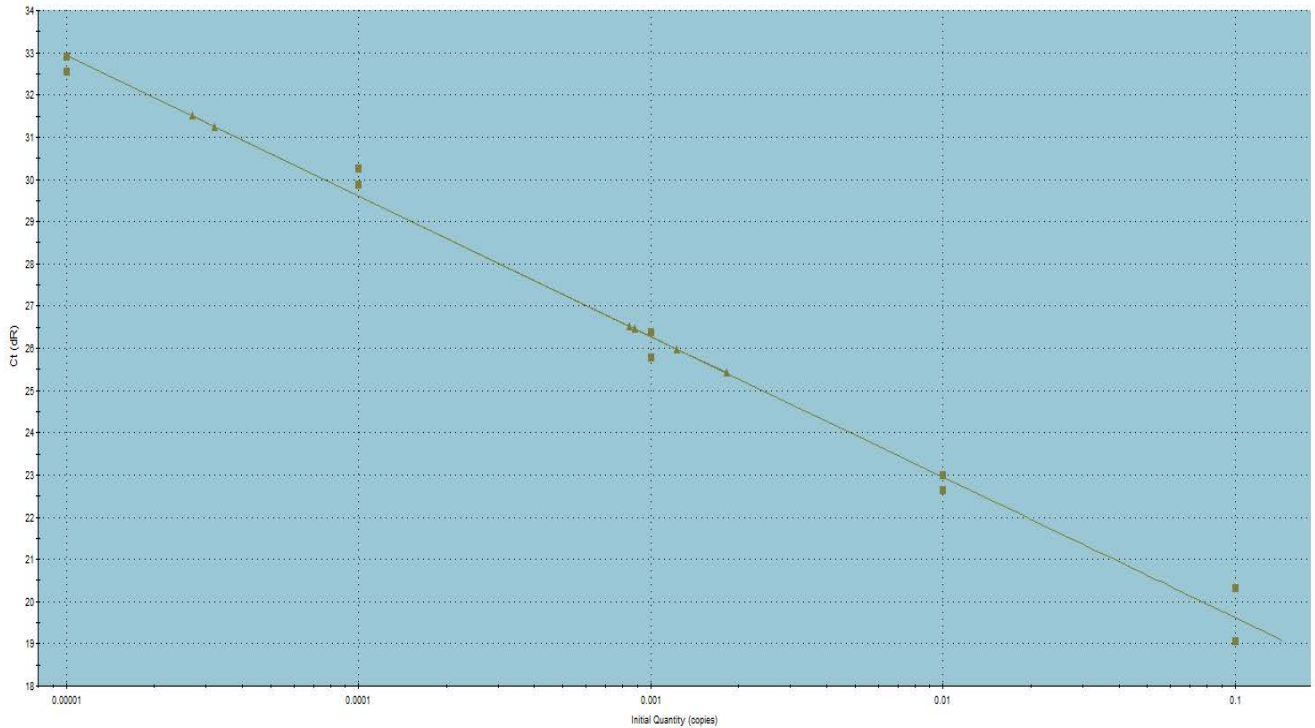


Figure 12 represents a standard curve generated by serially diluted sheared chromatin. Unknown CHIP DNA is represented by ▲ and standards are represented by ■. A RSq value of 0.992 was noted along with 99.6% efficiency. Initial quantity (copies) vs Ct(dR) was used to plot the standard curve

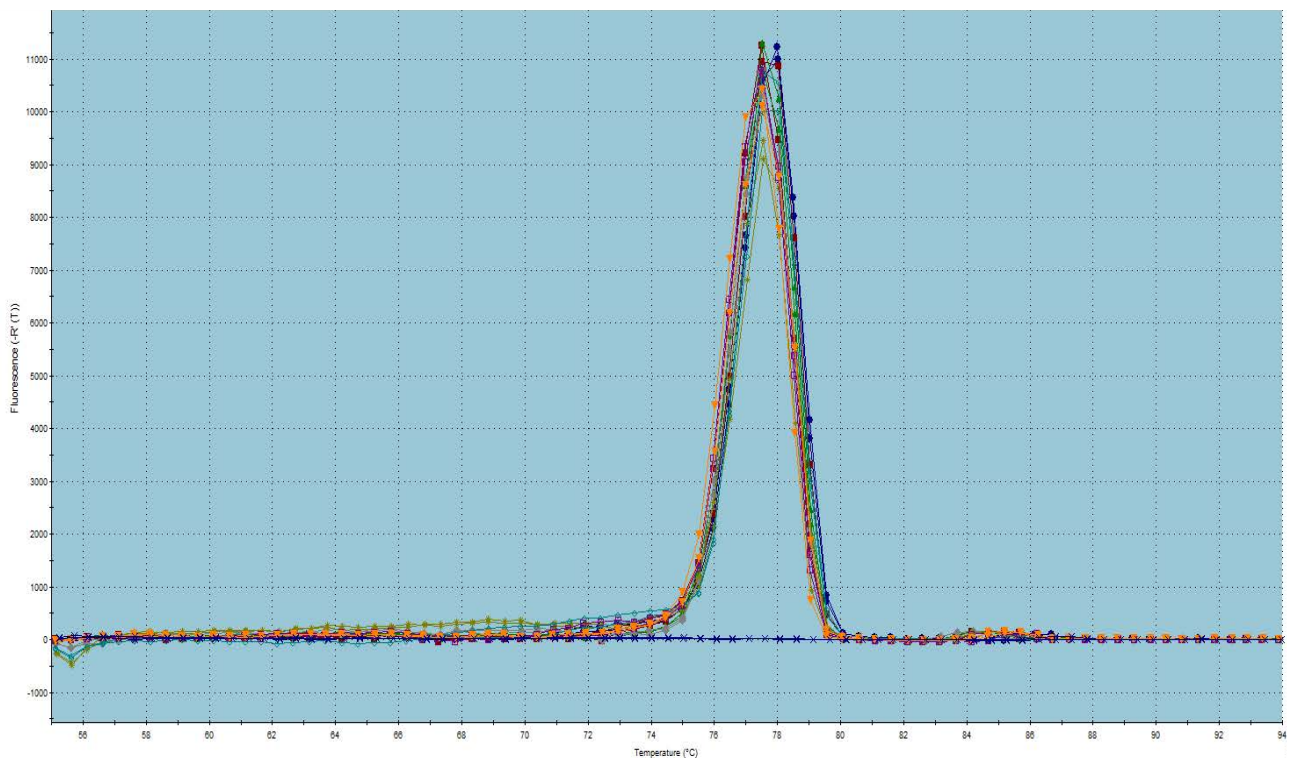


Figure 13 represents a dissociation curve with products aligning as a single peak. Temperature (°C) versus fluorescence (-R¹ CT) value was used to plot the graph

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.

Primary human tonsil keratinocytes with episomal HPV16 genomes

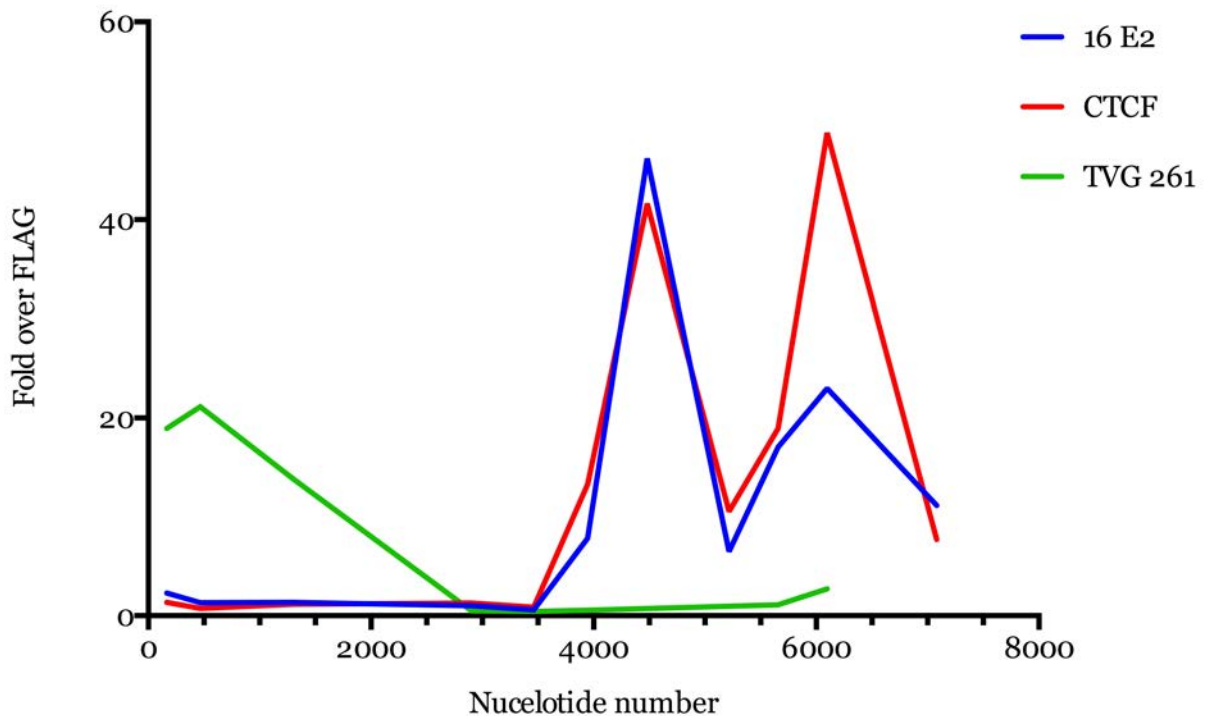


Figure 14 Real-time PCR analysis of immunoprecipitated chromatin derived from primary human tonsil keratinocytes with episomal HPV 16 genomes. Chromatin was sheared by enzymatic digestion. Samples were immunoprecipitated with HPV 16 Sheep E2 antibody (Parish Lab), HPV16 E2 TVG 261 antibody and CTCF antibody. 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

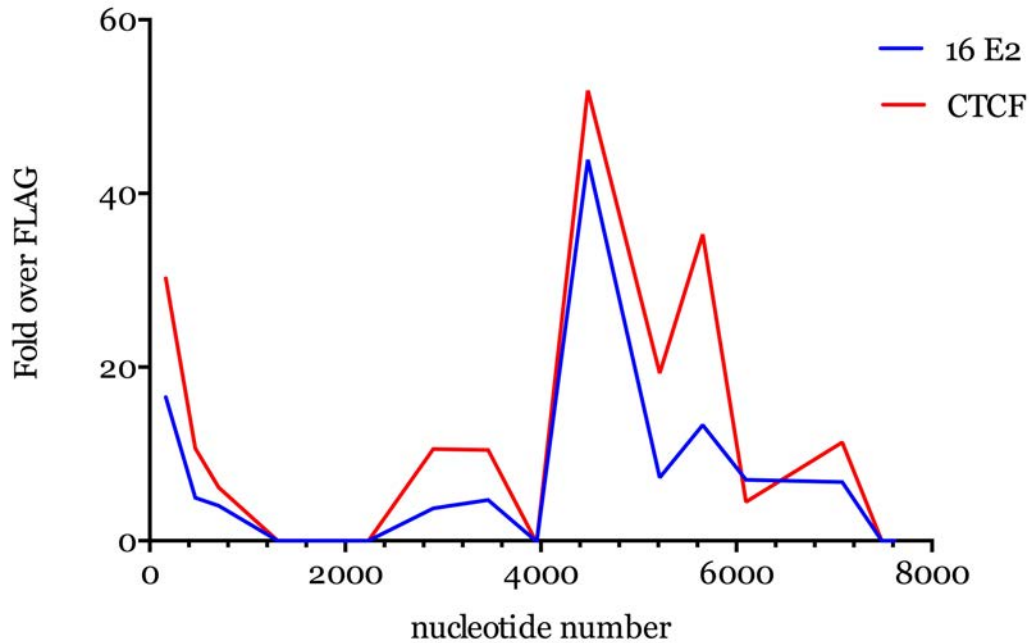


Figure 15 represents the real-time PCR analysis of the chromatin derived from primary human tonsil keratinocytes with episomal HPV 16 genomes. Chromatin was sheared with enzymatic cocktail. Samples were immunoprecipitated with HPV 16 Sheep E2 antibody (Parish Lab), TVG 261 antibody and CTCF antibody. 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 bioanalyser 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.

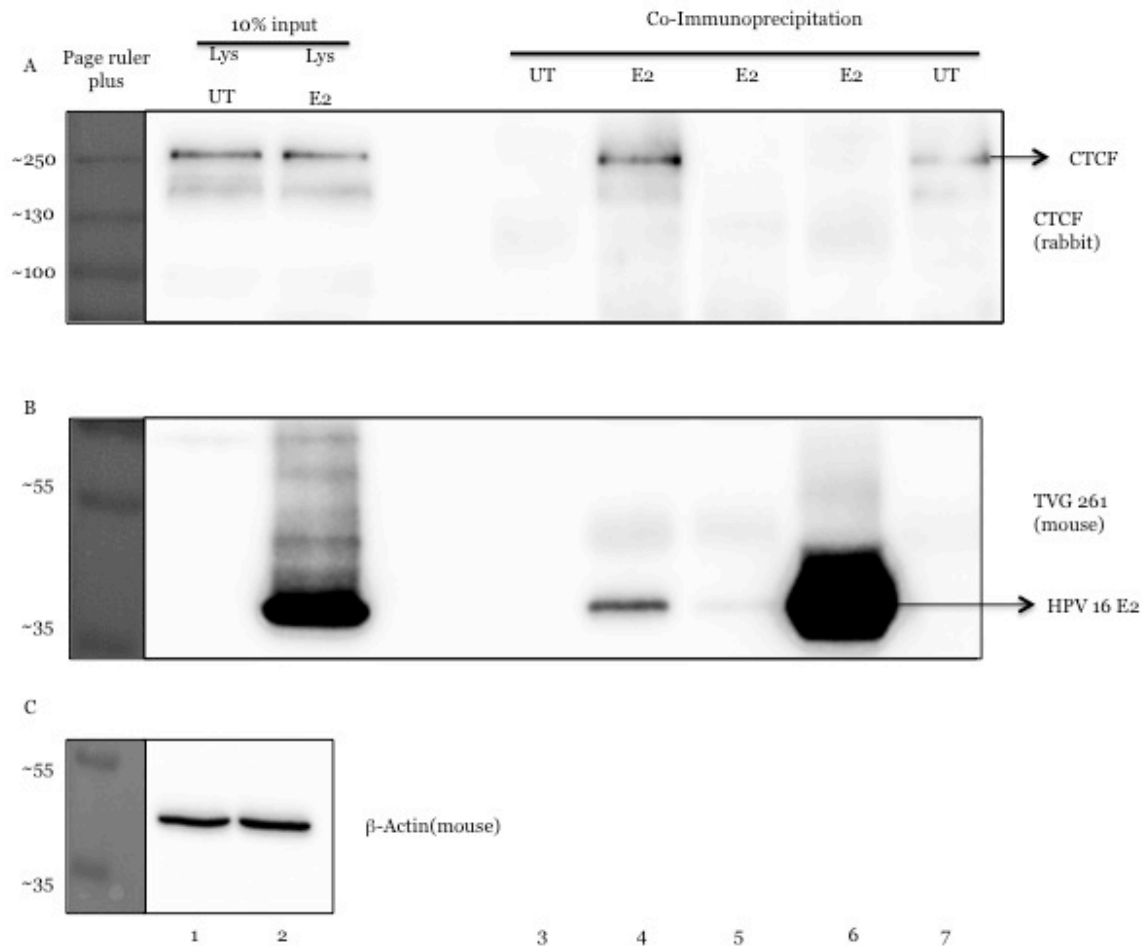


Figure 16 Co-immunoprecipitation of CTCF with HPV16 E2. 10% input samples are on the left (lane 1 and 2) and the co-immunoprecipitated samples on the right (lane 3-7). The arrows indicate the stated protein. Membrane A was incubated with CTCF antibody, membrane B was incubated with HPV 16 E2 sheep antibody 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

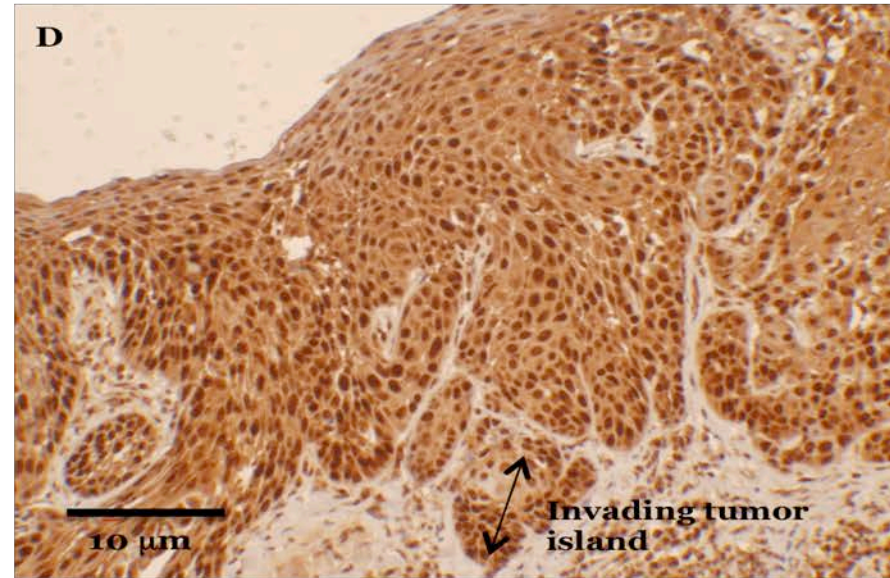
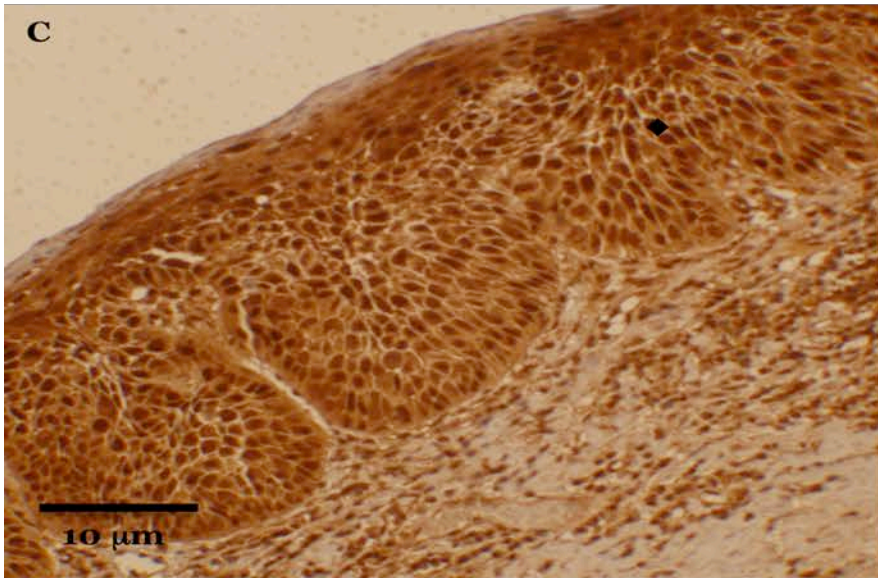
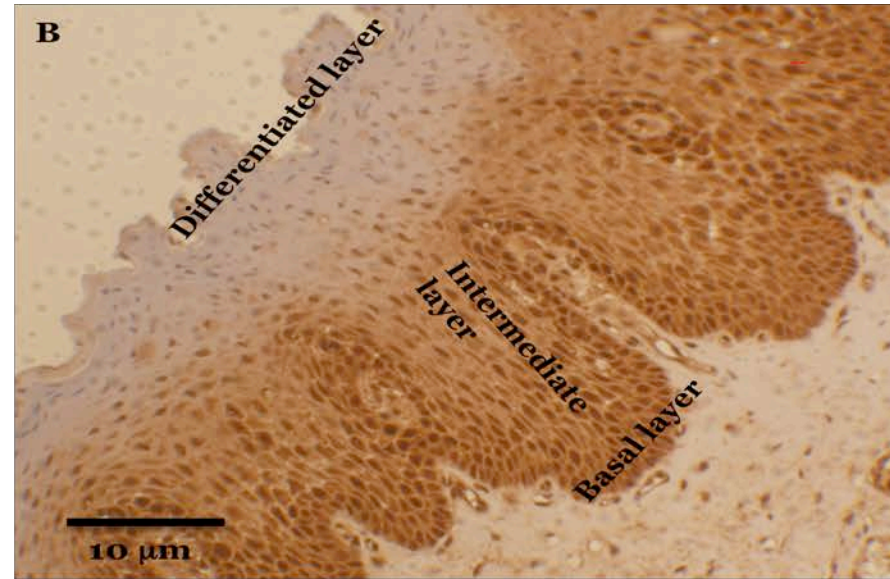
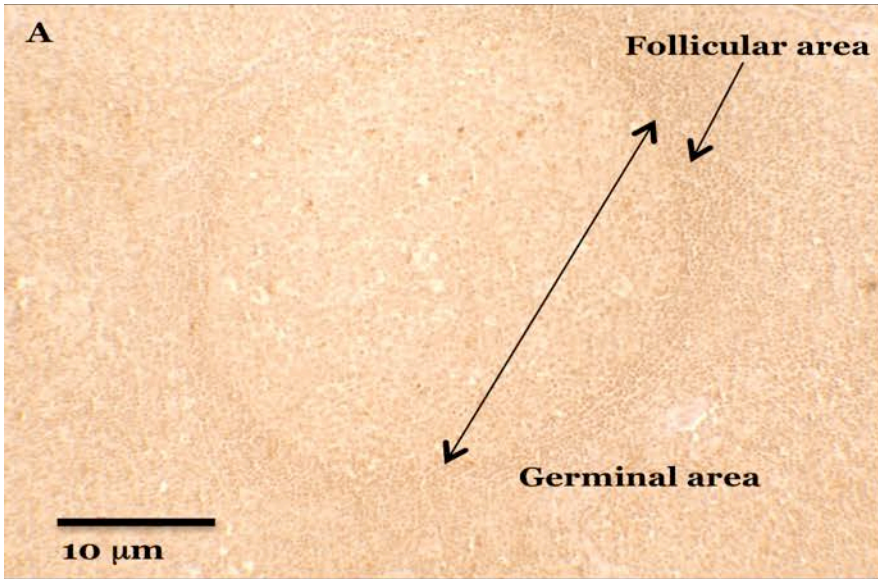
So far I have been investigating where in the genome CTCF and HPV 16 E2 are binding using an *in vitro* model system, however expression of CTCF *in vivo* may provide valuable information on disease progression and development. In order to investigate the expression of CTCF in physiologically normal 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.

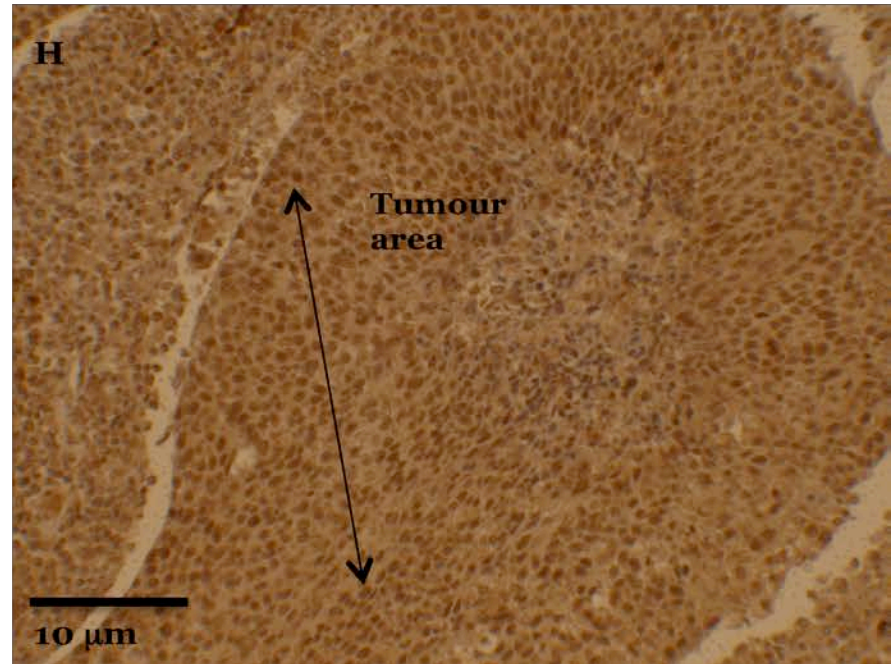
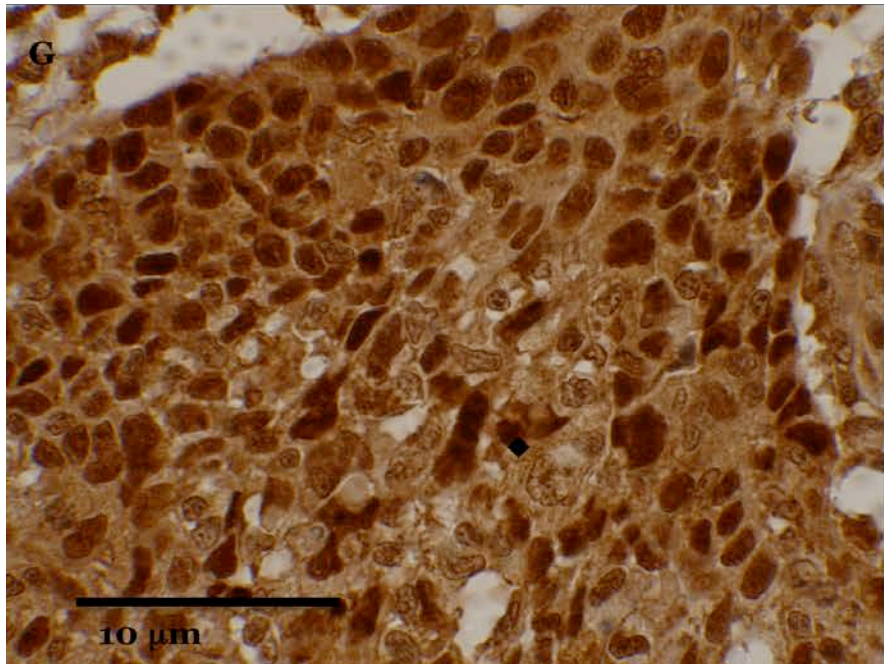
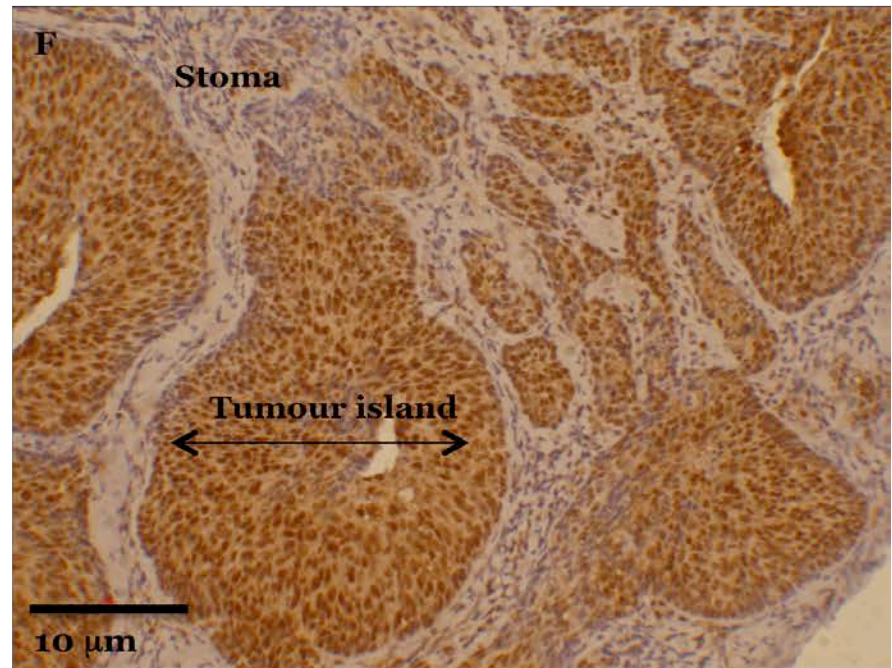
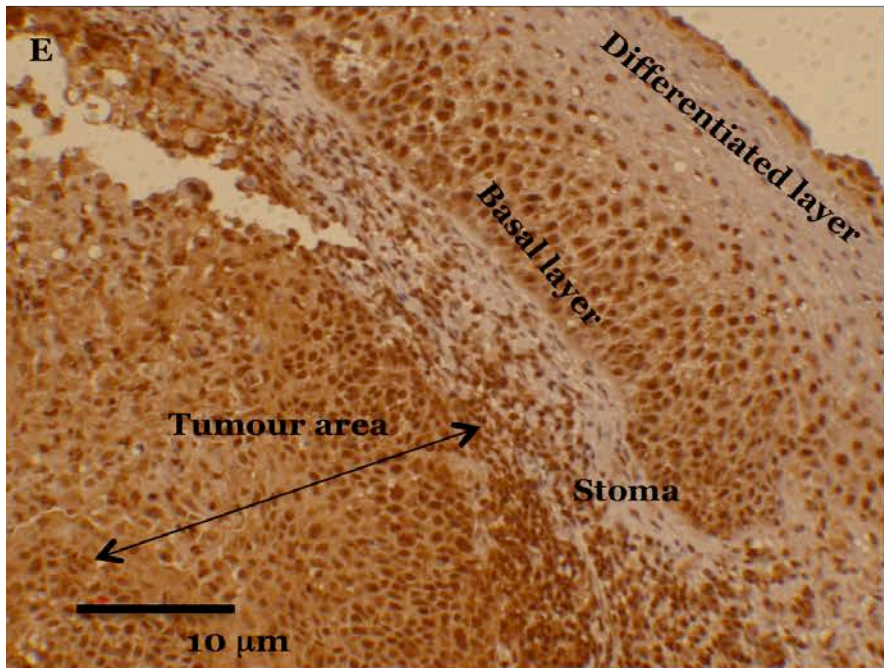
In the normal tonsil section, CTCF expression was detected in majority of the areas. However expression of CTCF in the germinal and follicular area were stronger compared with other areas of the normal tonsil section (Fig 17 A). Positive staining with CTCF was observed for the reticular crypt epithelium 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.





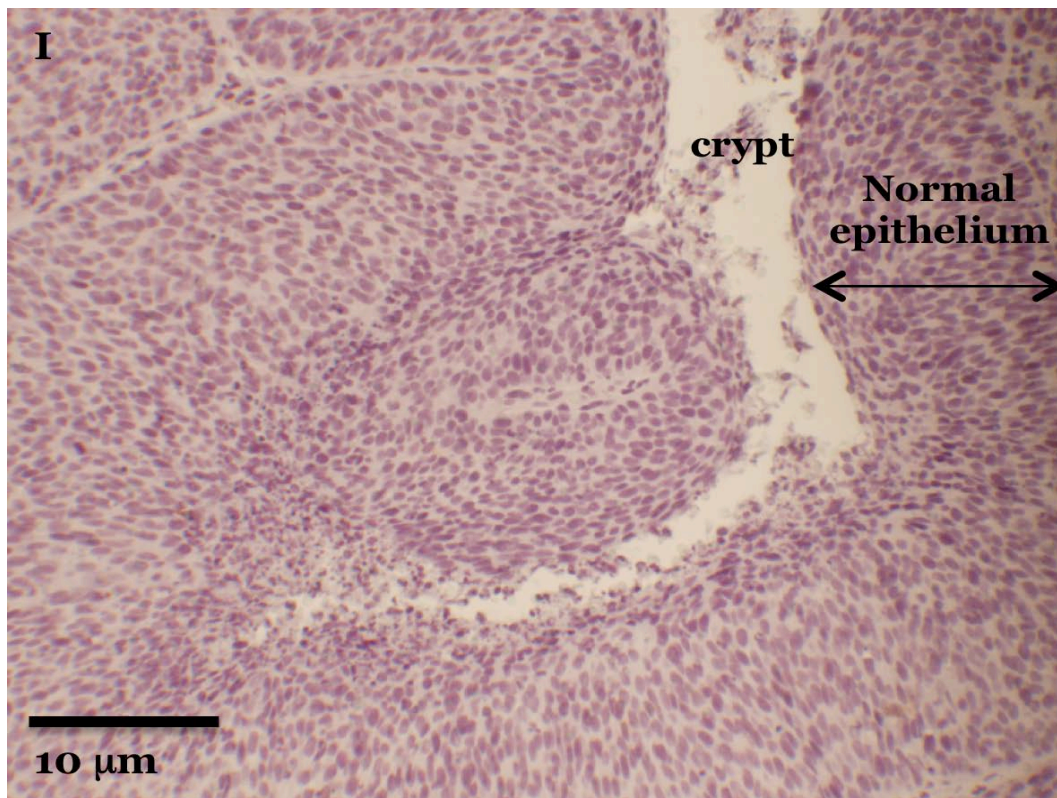


Figure 17 Representative photomicrographs showing of immunohistochemical analysis of CTCF expression. (A) germinal centre of the normal tonsil exhibits a strong CTCF expression in comparison to other areas within the normal tonsil section. (B) surface epithelium of the normal tonsil showing strong CTCF 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^{149,150}. HPV16 E2 is a sequence specific protein, which binds to its consensus sequence and initiates transcription and replication of the viral genome¹. In high-risk HPV genomes, there are four highly conserved E2 binding sites located in the viral locus control region¹⁴⁸. 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)₇GT is present at positions (35, 59, 499, 3786, 4591, 5117, 5872, 7019, 7451 and 7858) and AC(N)₇GGT 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 ¹⁴⁸. 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 ⁷² and methylation of the E2 binding site abrogates E2 binding and transcription regulation ^{151,152}. It has been reported that

there are five CpG islands within the HPV16 LCR 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 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 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 *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 ¹⁵⁶. pA controls the expression of proteins by adding a poly (A) tail on the mRNA, which acts as a signal to RNA polymerase II indicating termination of transcription¹⁵⁷. 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 ¹⁵⁶. Further analysis suggested the E2 was recruited to 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 sites 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 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 *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 ¹⁵⁹. 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 ¹⁵⁹. 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.

APPENDIX

Primer Pair	Sequence	Efficiency
L2 Fw	CAGGGTCGGGTACAGGCGGA	1.9347
L2 Rv	GGATCGGAAGGGCCCCACAGGA	
L2 CTCTbs Fw	AGGCGTACTGGCATTAGGTACAGT	1.7444
L2 CTCTbs Rv	AGGTAAGGCTGCATGTGAAGTGGT	
L2/L1 Fw	TGGCTGCCTAGAGGCCACTGT	1.9
L2/L1 Rv	TGCGTGCAACATATTCATCCGTGC	
L1 CTCFbs Fw	TGCAGCAAATGCAGGTGTGGAT	1.8647
L1 CTCFbs Rv	TGGGGATCCTTTGCCCCAGTGT	
L1 Fw	ACAAGCAGGATTGAAGGCCAAACCA	1.9379
L1 Rv	AGAGGTAGATGAGGTGGTGGGTGT	
5'URR-Enh Fw	TTTGTAGCGCCAGCGGCCATTT	1.8501
5'URR-Enh Rv	GCATGGCAAGCAGGAAACGTACAA	
Enh Fw	CCAAATCCCTGTTTTCTGA	1.6585
Enh Rv	CGTTGGCGCATAGTGATTTA	
Earlyprom Fw	GCAAACCGTTTTGGGTTACA	1.5128
Earlyprom rv	ACTAACCGTTTTCGGTTCAA	
E6 111-223 Fw	AGGACCCACAGGAGCGACCC	1.8695
E6 111-223 Rv	ACGTCGCAGTAACTGTTGCTTGCA	
E6 427-506 Fw	GCCACTGTGTCCTGAAGAAAAGCA	1.9332
E6 427-506 Rv	GACCGGTCCACCGACCCCTT	
Lateprom Fw	GACAGCTCAGAGGAGGAGGA	1.9004
Lateprom Rv	GCACAACCGAAGCGTAGAGT	
E1 1250-1368 Fw	GCGAAGACAGCGGGTATGGCA	1.9154
E1 1250-1368 rv	GCAACCACCCCACTTCCACC	
E1 2158-2316 Fw	AGGGTAGATGATGGAGGTGATTGG	1.8461
E1 2158-2316 Rv	GATTTACCTGTGTTAGCTGCACCA	
E2 (CTCF) Fw	GGAAACACATGCGCCTAGAATGTGC	1.8923
E2 (CTCF) Rv	TGATACAGCCAGTGTGGCACC	
E4 Fw	CACTCCGCCGCGACCCATAC	1.9326
E4 Rv	GGTGTGGCAGGGGTTTCCGG	
E5 Fw	ACGTCCGCTGCTTTTGTCTGTGT	1.9129

E5 Rv	ACCTAAACGCAGAGGCTGCTGT
-------	------------------------

Table 4 represents the primers used during qPCR, which cover the entire HPV16 genome. The primers were diluted to a final concentration of 10 pmolml⁻¹

NEB #	Product	Index Primer Sequence	Expected Index Primer Sequence Read
#E7339A: 0.010 ml #E7339AA: 0.020 ml	NEBNext Index 1 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGAT- CGTGAT GTGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	ATCACG
#E7340A: 0.010 ml #E7340AA: 0.020 ml	NEBNext Index 2 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGAT- ACATCGG TGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	CGATGT
#E7341A: 0.010 ml #E7341AA: 0.020 ml	NEBNext Index 3 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGA- TGCCTAA GTGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	TTAGGC
#E7342A: 0.010 ml #E7342AA: 0.020 ml	NEBNext Index 4 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGAT- TGGTCAG TGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	TGACCA
#E7343A: 0.010 ml #E7343AA: 0.020 ml	NEBNext Index 5 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGAT- CACTGTG TGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	ACAGTG
#E7344A: 0.010 ml #E7344AA: 0.020 ml	NEBNext Index 6 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGA- TAITGGC GTGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	GCCAAT
#E7345A: 0.010 ml #E7345AA: 0.020 ml	NEBNext Index 7 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGA- TGATCTG GTGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	CAGATC
#E7346A: 0.010 ml #E7346AA: 0.020 ml	NEBNext Index 8 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGA- TTCAGT GTGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	ACTTGA
#E7347A: 0.010 ml #E7347AA: 0.020 ml	NEBNext Index 9 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGA- TCTGATC GTGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	GATCAG
#E7348A: 0.010 ml #E7348AA: 0.020 ml	NEBNext Index 10 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGAT- AAGCTA GTGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	TAGCTT
#E7349A: 0.010 ml #E7349AA: 0.020 ml	NEBNext Index 11 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGA- TGTAGCC GTGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	GGCTAC
#E7331A: 0.010 ml #E7331AA: 0.020 ml	NEBNext Index 12 Primer for Illumina	5'-CAAGCAGAAGACGGCATAACGAGAT- TACAAG GTGACTGGAGTTCAGACGT- GTGCTCTCCGATC-s-T-3'	CTTGTA

Table 5 represent the lists of NEBNext Index primers from which Index primer number 4, 6 and 12 were used to construct the CHIP-SEQ

REFERENCES

1. Dell, G., Wilkinson, K., Tranter, R., Parish J., Brady, L. and Gaston, K. Comparison of the Structure and DNA-binding Properties of the E2 Proteins from an Oncogenic and a Non-oncogenic Human Papillomavirus. *Journal of Molecular Biology* (2003) **334**, 979–991.
2. Zheng, Z. M. & Baker, C. C. Papillomavirus genome structure, expression, and post-transcriptional regulation. *Frontiers in Bioscience* (2006) **11**:2286-2302
3. Klymenko, T. & Graham, S. V. Human papillomavirus gene expression is controlled by host cell splicing factors. *Biochem. Soc. Trans.* (2012) **40**, 773–777.
4. Doorbar, J. Quint, W., Banks, L., Bravo, I., Stoler, M., Broker, T. and Stanley, M. The biology and life-cycle of human papillomaviruses. *Vaccine* (2012)**30 Suppl 5**, F55–70.
5. Doorbar, J. Latent papillomavirus infections and their regulation. *Current Opinion in Virology* (2013) **3**, 416–421.
6. Smal, C., Welzler, D.E., Dantur, KI., Chemes, LB., Garcia-alai, MM., Dellarole M., Alonso LG., Gaston K. and de Prat-Gay, G. The human papillomavirus E7-E2 interaction mechanism in vitro reveals a finely tuned system for modulating available E7 and E2 proteins. *Biochemistry* (2009) **48**, 11939–11949.
7. Cobo, F. Human Papillomavirus Infections. Cambridge Elsevier Science and Technology (2012).
8. Doorbar, J. The papillomavirus life cycle. *Journal of Clinical Virology.* (2005) **32** Suppl **1**, S7–15.
9. Gottschling, M., Göker, M. & Köhler, A.. Lehman MD., Stockfleth, E and Nindl, I. Cutaneotropic Human β -/ γ -Papillomaviruses Are Rarely Shared between Family Members. *Journal of Investigative Dermatology* **129** (2009), 2427-2434.
10. Sánchez, I. E., Dellarole, M., Gaston, K. & de Prat-Gay, G. Comprehensive comparison of the interaction of the E2 master regulator with its cognate target DNA sites in 73 human papillomavirus types by sequence statistics. *Nucleic Acids Research* (2008) **36**, 756–769.
11. Middleton, K. Peh, W., Sourther, S., Griffin, H., Sotlar, K., Nakahara, T., El-Sherif, A., Morris, L., Seth, R., Hibma, M., Jenkins, D., Lambert, P., Coleman, N. and Doorbar, J. Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers. *Journal of Virology* (2003) **77**, 10186–10201.
12. Muñoz, N. Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV *et al* Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N. Engl. J. Med.* (2003) **348**, 518–527.
13. Clifford, G. M., Smith, J. S., Plummer, M. & Muñoz, N. Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *British journal of Cancer.* (2003) **88**, 63-73.
14. Parkin, D. M. Cancers attributable to infection in the UK in 2010. *British Journal*

- of Cancer* (2011) **105 Suppl 2**, S49–56.
15. Xue, Y, Bellanger S, Zhang W, Lim D, Low J, Lunny D *et al.* HPV16 E2 is an immediate early marker of viral infection, preceding E7 expression in precursor structures of cervical carcinoma. *Cancer Res* (2010) **70**, 5316–5325.
 16. Vizcaino, A. P., Moreno, V. & Bosch, F. X. International trends in incidence of cervical cancer: II. Squamous-cell carcinoma. *International journal Cancer* (2000) **86**, 429-435
 17. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer* (2010) **127**, 2893–2917.
 18. Parkin, D. M. & Bray, F. Chapter 2: The burden of HPV-related cancers. *Vaccine* (2006) **24 Suppl 3**, S3–11–25.
 19. Bruni L, Diaz M, Castellsague X, Ferrer E, Bosch FX, de Sanjose S.. Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. *J. Infect. Dis.* (2010) **202**, 1789–1799.
 20. Peirson, L, Fitzpatrick-Lewis, D., Ciliska, D. & Warren, R. Screening for cervical cancer: a systematic review and meta-analysis. *Systematic Reviews* (2013) **2**, 1–1.
 21. Arney, A. & Bennett, K. M. Molecular Diagnostics of Human Papillomavirus. *Laboratory Medicine* (2010) **41**, 523–530.
 22. Fernandes, J. V. & Araújo, J. Biology and natural history of human papillomavirus infection. *Open Access Journal of Clinical Trials* (2013) **5**, 1-12.
 23. Schiffman, M., Clifford, G. & Buonaguro, F. M. Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline. *Infect Agents Cancer* (2009) **4**, 8.
 24. Einstein, M. H. *et al.* Comparison of the immunogenicity and safety of Cervarix™ and Gardasil® human papillomavirus (HPV) cervical cancer vaccines in healthy women aged 18–45 years. *Human Vaccines* (2009) **5**, 705.
 25. Gellin, B., Modlin, J. F., Barr, E. & Tamms, G. Quadrivalent Human Papillomavirus Vaccine. *Clinical Infectious Diseases* (2007) **45**, 609–617.
 26. De Carvalho, N., Teixeira, J., Roteli-Martins, C. M. & Naud, P. Sustained efficacy and immunogenicity of the HPV-16/18 AS04-adjuvanted vaccine up to 7.3 years in young adult women. *Vaccine* (2010) **28**, 6247-6255.
 27. David, M. P., Van Herck, K., Hardt, K., Tibaldi, F. & Dubin, G. Long-term persistence of anti-HPV-16 and-18 antibodies induced by vaccination with the AS04-adjuvanted cervical cancer vaccine: modeling of sustained antibody responses. *Gynecologic Oncology* (2009)**115**, S1-6.
 28. Villa, L. L., Costa, R., Petta, C. A. & Andrade, R. P. High sustained efficacy of a prophylactic quadrivalent human papillomavirus types 6/11/16/18 L1 virus-like particle vaccine through 5 years of follow-up. *British journal of Cancer* (2006) **95**, 1459-1466.
 29. Hildesheim A, Herrero R, Wacholder S, Rodriguez AC, Solomon D, Bratti MC *et al.* Effect of human papillomavirus 16/18 L1 virus like particle vaccine among young women with preexisting infection: a randomized trial. *JAMA* (2007) **298**, 743–753.

30. Rogers, A., Waltke, M. & Angeletti, P. C. Evolutionary variation of papillomavirus E2 protein and E2 binding sites. *Virology Journal* (2011) **8**, 379.
31. Bodily, J. & Laimins, L. A. Persistence of human papillomavirus infection: keys to malignant progression. *Trends Microbiol.* (2011) **19**, 33–39.
32. Schiller, J. T., Day, P. M. & Kines, R. C. Current understanding of the mechanism of HPV infection. *Gynecol. Oncol.* (2010) **118**, S12–7.
33. Moody, C. A. & Laimins, L. A. Human papillomavirus oncoproteins: pathways to transformation. *Nature Publishing Group* (2010) **10**, 550–560.
34. Stenlund, A. Initiation of DNA replication: lessons from viral initiator proteins. *Nat. Rev. Mol. Cell Biol.* (2003)**4**, 777–785.
35. Steger, G. & Corbach, S. Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein. *Journal of Virology* (1997) **71**, 50–58.
36. Munger K, Baldwin A, Edwards KM, Hayakawa H, Nguyen CL, Owens M *et al.* Mechanisms of Human Papillomavirus-Induced Oncogenesis. *Journal of Virology* (2004) **78**, 11451–11460.
37. Bouvard, V., Storey, A., Pim, D. & Banks, L. Characterization of the human papillomavirus E2 protein: evidence of trans-activation and trans-repression in cervical keratinocytes. *The EMBO Journal* (1994) **13**, 5451–5459.
38. Stanley, M. Immunobiology of HPV and HPV vaccines. *Gynecol. Oncol.* (2008) **109**, S15–21.
39. Fehrman, F., Klumpp, D. J. & Laimins, L. A. Human Papillomavirus Type 31 E5 Protein Supports Cell Cycle Progression and Activates Late Viral Functions upon Epithelial Differentiation. *Journal of Virology* (2003) **77**, 2819–2831.
40. Spink, K. M. & Laimins, L. A. Induction of the human papillomavirus type 31 late promoter requires differentiation but not DNA amplification. *Journal of Virology* (2005) **79**, 4918–4926.
41. Zhang, B., Chen, W. & Roman, A. The E7 proteins of low- and high-risk human papillomaviruses share the ability to target the pRB family member p130 for degradation. *Proceedings of the National Academy of Sciences* (2006) **103**, 437–442.
42. Barrow-Laing, L., Chen, W. & Roman, A. Low- and high-risk human papillomavirus E7 proteins regulate p130 differently. *Virology* (2010) **400**, 233–239.
43. Cole, S. T. & Danos, O. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome: phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. *Journal of Molecular Biology* (1987) **193**, 599–608.
44. Barbosa, M. S., Lowy, D. R. & Schiller, J. T. Papillomavirus polypeptides E6 and E7 are zinc-binding proteins. *Journal of Virology* (1989) **63**, 1404–1407.
45. Slee, E. A., O'Connor, D. J. & Lu, X. To die or not to die: how does p53 decide? *Oncogene* (2004) **23**, 2809–2818.
46. Oren, M. Decision making by p53: life, death and cancer. *Cell Death & Differentiation* (2003) **10**, 431–442.
47. Fu L, Van Doorslaer K, Chen Z, Ristriani T, Masson M, Trave G, *et al.* Degradation of p53 by human Alpha-papillomavirus E6 proteins shows a

- stronger correlation with phylogeny than oncogenicity. *PLoS ONE* (2010) **5**, 9.
48. Pim, D. & Banks, L. Interaction of viral oncoproteins with cellular target molecules: infection with high-risk vs low-risk human papillomaviruses. *APMIS* (2010) **118**, 471–493.
 49. Huibregtse, J. M., Scheffner, M. & Howley, P. M. A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *The EMBO Journal* (1991) **10**, 4129–4135.
 50. Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J. & Howley, P. M. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* (1990) **63**, 1129–1136.
 51. Gewin, L. & Galloway, D. A. E box-dependent activation of telomerase by human papillomavirus type 16 E6 does not require induction of c-myc. *Journal of Virology* (2001) **75**, 7198–7201.
 52. Galloway DA, Gewin LC, Myers H, Luo W, Grandori C, Katzenellenbogen RA *et al.* Regulation of telomerase by human papillomaviruses. *Cold Spring Harb. Symp. Quant. Biol.* (2005) **70**, 209–215.
 53. Kranjec, C. & Banks, L. A Systematic Analysis of Human Papillomavirus (HPV) E6 PDZ Substrates Identifies MAGI-1 as a Major Target of HPV Type 16 (HPV-16) and HPV-18 Whose Loss Accompanies Disruption of Tight Junctions. *Journal of Virology* (2011) **85**, 1757–1764.
 54. Javier, R. T. Cell polarity proteins: common targets for tumorigenic human viruses. *Oncogene* (2008) **27**, 7031–7046.
 55. Dyson, N., Guida, P., Munger, K. & Harlow, E. Homologous sequences in adenovirus E1A and human papillomavirus E7 proteins mediate interaction with the same set of cellular proteins. *Journal of Virology* (1992) **66**, 6893–6902.
 56. Barbosa MS, Edmonds C, Fisher C, Schiller JT, Lowy DR, Vousden KH. The region of the HPV E7 oncoprotein homologous to adenovirus E1a and Sv40 large T antigen contains separate domains for Rb binding and casein kinase II phosphorylation. *The EMBO Journal* (1990) **9**, 153–160.
 57. Chellappan S, Kraus VB, Kroger B, Munger K, Howley PM, Phelps WC. *et al.* Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product. *Proc. Natl. Acad. Sci. U.S.A.* (1992) **89**, 4549–4553.
 58. Boyer, S. N., Wazer, D. E. & Band, V. E7 Protein of Human Papilloma Virus-16 Induces Degradation of Retinoblastoma Protein through the Ubiquitin-Proteasome Pathway. *Cancer Res* (1996) **56**, 4620–4624.
 59. Berezutskaya, E., Yu, B. & Morozov, A. Differential regulation of the pocket domains of the retinoblastoma family proteins by the HPV16 E7 oncoprotein. *Cell growth & Differentiation* (1997) **8**, 1277–1286.
 60. Sanchez-Perez, A. M., Soriano, S., Clarke, A. R. & Gaston, K. Disruption of the human papillomavirus type 16 E2 gene protects cervical carcinoma cells from E2F-induced apoptosis. *J. Gen. Virol.* (1997) **78**, 3009–3018.
 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-

- induced apoptosis. *Virology* (1997) **239**, 97–107.
62. Burkhardt, A., Willingham, M., Gay, C. & Jeang, K. T. The E5 oncoprotein of bovine papillomavirus is oriented asymmetrically in Golgi and plasma membranes. *Virology* (1989) **170**, 334–339.
 63. Conrad, M., Bubb, V. J. & Schlegel, R. The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein. *Journal of Virology* (1993) **67**, 6170–6178.
 64. Venuti A, Paolini F, Nasir L, Corteggio A, Roperto S, Campo M. *et al.* Papillomavirus E5: the smallest oncoprotein with many functions. *Molecular Cancer* (2011) **10**, 140.
 65. Valle, G. F. & Banks, L. The human papillomavirus (HPV)-6 and HPV-16 E5 proteins co-operate with HPV-16 E7 in the transformation of primary rodent cells. *J. Gen. Virol.* (1995) **76**, 1239–1245.
 66. Bouvard, V., Matlashewski, G., Gu, Z. M., Storey, A. & Banks, L. The human papillomavirus type 16 E5 gene cooperates with the E7 gene to stimulate proliferation of primary cells and increases viral gene expression. *Virology* (1994) **203**, 73–80.
 67. Straight, S. W., Hinkle, P. M., Jewers, R. J. & McCance, D. J. The E5 oncoprotein of human papillomavirus type 16 transforms fibroblasts and effects the downregulation of the epidermal growth factor receptor in keratinocytes. *Journal of Virology* (1993) **67**, 4521–4532.
 68. Crusius, K., Auvinen, E. & Alonso, A. Enhancement of EGF-and PMA-mediated MAP kinase activation in cells expressing the human papillomavirus type 16 E5 protein. *Oncogene* (1997) **15**, 1437–1444.
 69. Hegde, R. The papillomavirus E2 proteins: structure, function, and biology. *Annual Reviews of Biophysics and Biomolecular Structure* (2002) **31**, 343-360.
 70. Vosa L, Sudakov A, Remm M, Ustav M, Kurg R Identification and analysis of papillomavirus E2 protein binding sites in the human genome. *Journal of Virology* (2012) **86**, 348–357.
 71. Giri, I. and Yaniv, M. Structural and mutational analysis of E2 trans-activating proteins of papillomaviruses reveals three distinct functional domains. *EMBO* (1988) **7**, 2823–2829.
 72. McBride, A. A. The papillomavirus E2 proteins. *Virology* (2013) **445**, 57–79.
 73. Kurg, R., Tekkel, H., Abroi, A. and Ustav, M. Characterization of the Functional Activities of the Bovine Papillomavirus Type 1 E2 Protein Single-Chain Heterodimers. *Journal of Virology* (2006) **80**, 11218–11225.
 74. Kurg, R., Uusen, P., Vosa, L. and Ustav, M. Human papillomavirus E2 protein with single activation domain initiates HPV18 genome replication, but is not sufficient for long-term maintenance of virus genome. *Virology* (2010) **408**, 159–166.
 75. Nadra, A., Elsie, T., Mok, Y-K., Fabio, C.L., Almeida, M., Bycroft, M. Paci, de Prat-Gay, G. and Cicero, D.O. Letter to Editor: Solution structure of the HPV-16 E2 DNA binding domain, a transcriptional regulator with a dimeric β -barrel fold. *Journal of Biomolecular NMR* (2004) **30**, 211–214.
 76. Antson AA, Burns JE, Moroz OV, Scott DJ, Sanders CM, Bronstein IB, *et al.* Structure of the intact transactivation domain of the human papillomavirus E2

- protein. *Nature* (2000) **403**, 805–809.
77. Pepinsky RB, Prakash SS, Corina K, Grossel MJ, Barsoum J, Androphy EJ Sequences flanking the core DNA-binding domain of bovine papillomavirus type 1 E2 contribute to DNA-binding function. (1997) **71**, 828–831.
 78. Soeda, E., Farren, M.C., Baker, C.C. and McBride, A.A. Repression of HPV16 early region transcription by the E2 protein. *Virology* (2006) **351**, 29–41.
 79. Bernard BA, Bailly C, Lenoir MC, Darmon M, Thierry F, Yaniv M. The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV18 regulatory region in human keratinocytes. *Journal of Virology* (1989) **63**, 4317–4324.
 80. Hou, S. Y., Wu, S. Y., Zhou, T., Thomas, M. C. & Chiang, C. M. Alleviation of human papillomavirus E2-mediated transcriptional repression via formation of a TATA binding protein (or TFIID)-TFIIB-RNA polymerase II-TFIIF preinitiation complex. *Mol. Cell. Biol.* (2000) **20**, 113–125.
 81. Lim, D. A., Gossen, M., Lehman, C. W. & Botchan, M. R. Competition for DNA binding sites between the short and long forms of E2 dimers underlies repression in bovine papillomavirus type 1 DNA replication control. *Journal of Virology* (1998) **72**, 1931–1940.
 82. Barsoum, J., Prakash, S. S., Han, P. & Androphy, E. J. Mechanism of action of the papillomavirus E2 repressor: repression in the absence of DNA binding. *Journal of Virology* (1992) **66**, 3941–3945.
 83. Mohr IJ, Clark R, Sun S, Androphy EJ, MacPherson P, Botchan MR. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. *Science* (1990) **250**, 1694–1699.
 84. Sanders, C. M. & Arne, S. Transcription factor-dependent loading of the E1 initiator reveals modular assembly of the papillomavirus origin melting complex. *Journal of Biological Chemistry* (2000) **275**, 3522–3534.
 85. Frattini, M. G. & Laimins, L. A. Binding of the human papillomavirus E1 origin-recognition protein is regulated through complex formation with the E2 enhancer-binding protein. *Proc. Natl. Acad. Sci. U.S.A.* (1994) **91**, 12398–12402.
 86. Sakakibara N, Chen D, Jang MK, Kang DW, Luecke HF, Wu S-Y. *et al.* Brd4 is displaced from HPV replication factories as they expand and amplify viral DNA. *PLoS Pathog* (2013) **9**, e1003777.
 87. Reinson T, Toots M, Kadaja M, Pipitch R, Allik M, Ustav E., *et al.* Engagement of the ATR-dependent DNA damage response at the human papillomavirus 18 replication centers during the initial amplification. *Journal of Virology* (2013) **87**, 951–964.
 88. Bastien, N. & McBride, A. A. Interaction of the papillomavirus E2 protein with mitotic chromosomes. *Virology* (2000) **270**, 124–134.
 89. Skiadopoulos, M. H. & McBride, A. A. Bovine papillomavirus type 1 genomes and the E2 transactivator protein are closely associated with mitotic chromatin. *Journal of Virology* (1998) **72**, 2079–2088.
 90. McPhillips, M.G., Oliveira, J.G., Spindler, J.E., Mitra, R. and McBride, A.A. Brd4 is required for e2-mediated transcriptional activation but not genome partitioning of all papillomaviruses. *Journal of Virology* (2006) **80**, 9530–9543.

91. Lusky, M. and Fontane, A. Formation of the complex of bovine papillomavirus E1 and E2 proteins is modulated by E2 phosphorylation and depends upon sequences within the carboxyl terminus of E1. *PNAS* (1991) **88**, 6363–6367.
92. Bergvall, M., Melendy, T. and Archambault, J. The E1 proteins. *Virology* (2013) **445**, 35–56.
93. Sverdrup, F. and Myers, G. The E1 proteins. *In: Human papillomaviruses* (Ed. G. Myers) (1997) 37-53.
94. Liu JS, Kuo SR, Makhov AM, Cyr DM, Griffith JD, Broker TR, et al. Human Hsp70 and Hsp40 chaperone proteins facilitate human papillomavirus-11 E1 protein binding to the origin and stimulate cell-free DNA replication. (1998) *J Virol* 273(46):30704-12.
95. Fouts, E., Yu, X., Egelman, E.H. and Botchan, M. Biochemical and electron microscopic image analysis of the hexameric E1 helicase. *Journal of biological Chemistry* (1999) **274**, 4447-4458
96. Doorbar, J. & Myers, G. The E4 protein In: Myers, G., Delius, H., Icenogel, J., Bernard, H.-U., Baker, C., Halpern, A. & Wheeler, C. editor. *Human Papillomaviruses*. Los Alamos, New Mexico: Los Alamos National Laboratory vol. III, pp. 58–80 (1996).
97. Chow, L. T., Nasser, M., Wolinsky, S. M. & Broker, T. R. Human papillomavirus types 6 and 11 mRNAs from genital condylomata acuminata. *Journal of Virology* (1987) **61**, 2581–2588.
98. Wang, X., Meyers, C., Wang, H.-K., Chow, L. T. & Zheng, Z.-M. Construction of a full transcription map of human papillomavirus type 18 during productive viral infection. *Journal of Virology* (2011) **85**, 8080–8092.
99. Nakahara, T., Peh, W.L., Doorbar, J., Lee, D. and Lambert P.F. Human papillomavirus type 16 E1^{E4} contributes to multiple facets of the papillomavirus life cycle. *Journal of Virology* (2005) **79**, 13150–13165.
100. Peh WL, Middleton K, Christensen N, Nicholls P, Egawa K, Sotlar K, et al. Life Cycle Heterogeneity in Animal Models of Human Papillomavirus-Associated Disease. *Journal of Virology* (2002)**76**, 10401–10416.
101. Neary, K., Horwitz, B. H. & DiMaio, D. Mutational analysis of open reading frame E4 of bovine papillomavirus type 1. *Journal of Virology* (1987) **61**, 1248–1252.
102. Khan, J., Davy, C.E., McIntosh P.B., Jackson, J., Hinz, S., Wang, Q. and Doorbar, J. Role of calpain in the formation of human papillomavirus type 16 E1^{E4} amyloid fibers and reorganization of the keratin network. *Journal of Virology* (2011) **85**, 9984–9997.
103. Davy, C. and Doorbar, J. G2/M cell cycle arrest in the life cycle of viruses. *Virology* (2007) **368**, 219–226.
104. Ozbun, M. A. & Meyers, C. Characterization of late gene transcripts expressed during vegetative replication of human papillomavirus type 31b. *Journal of Virology* (1997) **71**, 5161–5172.
105. Favre, M. Structural polypeptides of rabbit, bovine, and human papillomaviruses. *Journal of Virology* (1975) **15**, 1239–1247.
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

- life cycle. *Journal of Virology* (2005) **79**, 3938–3948.
107. Barouch, D. H. & Harrison, S. C. Interactions among the major and minor coat proteins of polyomavirus. *Journal of Virology* (1994) **68**, 3982–3989.
 108. Yang, R. Interaction of L2 with beta -Actin Directs Intracellular Transport of Papillomavirus and Infection. *Journal of Biological Chemistry* (2003)**278**, 12546–12553.
 109. Kämper, N, Day PM, Nowak T, Selinka HC, Florin L, Bolscher J, *et al.* A membrane-destabilizing peptide in capsid protein L2 is required for egress of papillomavirus genomes from endosomes. *Journal of Virology* (2006) **80**, 759–768.
 110. Paintsil, J., Müller, M., Picken, M., Gissmann, L. & Zhou, J. Carboxyl terminus of bovine papillomavirus type-1 L1 protein is not required for capsid formation. *Virology* (1996) **223**, 238–244.
 111. Richards, R. M., Lowy, D. R., Schiller, J. T. & Day, P. M. Cleavage of the papillomavirus minor capsid protein, L2, at a furin consensus site is necessary for infection. *Proc. Natl. Acad. Sci. U.S.A.* (2006)**103**, 1522–1527.
 112. Buck, C. B., Day, P. M. & Trus, B. L. The papillomavirus major capsid protein L1. *Virology* (2013) **445**, 169–174.
 113. Graham, S. V. Human papillomavirus: gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. *Future Microbiology* (2010) **5**, 1493–1506.
 114. Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* (2009) **19**, 24–32.
 115. Lobanenkov, V. V., Nicolas, R. H., Adler, V. V. & Paterson, H. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* (1990) **5**, 1743-1753.
 116. Baniahmad, A., Steiner, C., Köhne, A. C. & Renkawitz, R. Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. *Cell* (1990) **61**, 505-514
 117. Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, *et al.* CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol. Cell. Biol.* (1993) **13**, 7612–7624.
 118. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* (2000) **405**, 486–489.
 119. Bell, A. C., West, A. G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* (1999) **98**, 387-396.
 120. Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, *et al.* Position-effect protection and enhancer blocking by the chicken β -globin insulator are separable activities. *Proc. Natl. Acad. Sci. U.S.A.*(2002) **99**, 6883–6888.
 121. Xu J, Huo D, Chen Y, Nwachukwu C, Collins C, Rowell J, *et al.* CpG island

- methylation affects accessibility of the proximal BRCA1 promoter to transcription factors. *Breast Cancer Res Treat* (2010) **120**, 593–601.
122. Majumder, P., Gomez, J. A., Chadwick, B. P. & Boss, J. M. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J. Exp. Med.* (2008) **205**, 785–798.
 123. Lutz M, Burke LJ, Barreto G, Goeman F, Greb H, Arnold R *et al.* Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res.* (2000) **28**, 1707–1713.
 124. Renaud, S., Loukinov, D., Bosman, F. T., Lobanenko, V. & Benhattar, J. CTCF binds the proximal exonic region of hTERT and inhibits its transcription. *Nucleic Acids Res.* (2005) **33**, 6850–6860.
 125. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* (2009) **459**, 108–112.
 126. Moon H, Filippova G, Loukinov D, Pugacheva E, Chen Q, Smith ST, *et al.* CTCF is conserved from *Drosophila* to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO reports* (2005) **6**, 165–170.
 127. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, *et al.* CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* (2006) **20**, 2349–2354.
 128. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* (2008) **132**, 422–433.
 129. Williams, A. & Flavell, R. A. The role of CTCF in regulating nuclear organization. *J. Exp. Med.* (2008) **205**, 747–750.
 130. Fu, Y., Sinha, M., Peterson, C. L. & Weng, Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* (2008) **4**, e1000138.
 131. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* (2012) **22**, 1680–1688.
 132. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, *et al.* Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* (2012) **148**, 335–348.
 133. Essien K, Vigneau S, Apreleva S, Singh LN, Bartolomei MS, Hannonhalli S. CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. *Genome Biol.* (2009) **10**, R131.
 134. MacPherson, M. J., Beatty, L. G., Zhou, W., Du, M. & Sadowski, P. D. The CTCF insulator protein is posttranslationally modified by SUMO. *Mol. Cell. Biol.* (2009) **29**, 714–725.
 135. Martinez, S. R. & Miranda, J. L. CTCF terminal segments are unstructured. *Protein Sci.* (2010) **19**, 1110–1116.
 136. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* (2009) **137**, 1194–1211
 137. El-Kady, A. & Klenova, E. Regulation of the transcription factor, CTCF, by

- phosphorylation with protein kinase CK2. *FEBS Lett.* (2005) **579**, 1424–1434.
138. Klenova EM, Chernukhin IV, El-Kady A, Lee RE, Pugacheva EM, Loukinov DI, *et al.* Functional Phosphorylation Sites in the C-Terminal Region of the Multivalent Multifunctional Transcriptional Factor CTCF. *Mol. Cell. Biol.* (2001) **21**, 2221–2234.
 139. Aker, M., Bomsztyk, K. & Emery, D. W. Poly(ADP-ribose) polymerase-1 (PARP-1) contributes to the barrier function of a vertebrate chromatin insulator. *J. Biol. Chem.* (2010) **285**, 37589–37597.
 140. Nakahashi H, Kwon KR, Resch W, Vian L, Dose M, Stavreva D, *et al.* A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep* (2013) **3**, 1678–1689.
 141. Singh, P., Lee, D.-H. & Szabó, P. E. More than insulator: multiple roles of CTCF at the H19-Igf2 imprinted domain. *Front Genet* (2012) **3**, 214.
 142. Fedoriw, A. M., Stein, P., Svoboda, P., Schultz, R. M. & Bartolomei, M. S. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. *Science* (2004) **303**, 238-240.
 143. Ong, C.-T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nature Reviews Genetics* (2014) **15**, 234–246.
 144. Dunn, K. L., Zhao, H. & Davie, J. R. The insulator binding protein CTCF associates with the nuclear matrix. *Exp. Cell Res.* (2003) **228**, 218-223.
 145. Torrano V, Navascues J, Docquier F, Zhang R, Burke LJ, Chernukhin I, *et al.* Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism. *J. Cell. Sci.* (2006) **119**, 1746–1759.
 146. Feeney, K. M. & Parish, J. L. Targeting mitotic chromosomes: a conserved mechanism to ensure viral genome persistence. *Proc. Biol. Sci.* (2009) **276**, 1535–1544.
 147. Zlatanova, J. & Caiafa, P. CCCTC-binding factor: to loop or to bridge. *Cellular and molecular life sciences* (2009). **66** 1647-1660151.
 148. Bhattacharjee, B. & Sengupta, S. CpG methylation of HPV 16 LCR at E2 binding site proximal to P97 is associated with cervical cancer in presence of intact E2. *Virology* (2006) **354**, 280–285.
 149. Thierry, F. Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. *Virology* (2009) **384**, 375–379.
 150. Hamid, N. A., Brown, C. & Gaston, K. The regulation of cell proliferation by the papillomavirus early proteins. *Cellular and molecular life sciences.* (2009) **66**, 1700-1717.
 151. Thain A, Jenkins O, Clarke AR, Gaston K CpG methylation directly inhibits binding of the human papillomavirus type 16 E2 protein to specific DNA sequences. (1996) **70**, 7233–7235.
 152. Kim, K., Garner-Hamrick, P. A., Fisher, C. & Lee, D. Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection. *Journal of Virology.* (2003) **77**, 12450-12459.
 153. Jacquin E, Baraquin A, Ramanah R, Carcopino X, Morel A, Valmary-Degano S, *et al.* Methylation of Human Papillomavirus Type 16 CpG Sites at E2-

- 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.
154. Snellenberg S, Schutze DM, Claassen-Kramer D, Meijer CJ, Snijders PJ, Steenbergen RD. Methylation status of the E2 binding sites of HPV16 in cervical lesions determined with the Luminex® xMAP™ system. *Virology* (2012) **422**, 357–365.
 155. Münger, K. & Howley, P. M. Human papillomavirus immortalization and transformation functions. *Virus Research* (2002) **89**, 213–228.
 156. Johansson C, Somberg M, Li X, Backstrom Winqvist E, Fay J, Ryan F, *et al.* HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation. *The EMBO Journal* (2012) **31**, 3212–3227.
 157. Proudfoot, N. J. Ending the message: poly(A) signals then and now. *Genes Dev.* (2011) **25**, 1770–1782.
 158. Kim, S. S., Tam, J. K., Wang, A. F. & Hegde, R. S. The Structural Basis of DNA Target Discrimination by Papillomavirus E2 Proteins. *Journal of Biological Chemistry* (2000) **275**, 31245–31254.
 159. Rosa-Garrido M, Ceballos L, Alonso-Lecue P, Abraira C, Delgado MD, Gandarillas A. A cell cycle role for the epigenetic factor CTCF-L/BORIS. *PLoS ONE* (2012) **7**, e39371.
 160. Ohlsson R, Renkawitz R, Lobanenkov V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends in genetics : TIG.* (2001) **17** 520-527.